

Le développement, validation et évaluation de l'utilité clinique d'un modèle de
prédiction de mortalité à un an pour stimuler une discussion sur les objectifs de
soins chez les patients hospitalisés

par

Ryeyan Taseen

mémoire présenté au Département d'informatique
en vue de l'obtention du grade de maître ès sciences (M.Sc.)

FACULTÉ DES SCIENCES
UNIVERSITÉ DE SHERBROOKE

Sherbrooke, Québec, Canada, 22 février 2021

Le 22 février 2021

Le jury a accepté le mémoire de Monsieur Ryeyan Taseen dans sa version finale.

Membres du jury

Professeur Jean-François Ethier
Directeur de recherche
Département de médecine

Professeur Luc Lavoie
Codirecteur de recherche
Département d'informatique

Professeur Félix Camirand Lemyre
Membre interne
Département de mathématiques

Professeur Martin Vallières
Président-rapporteur
Département d'informatique

Sommaire

Les discussions sur les objectifs de soins (DOS) n'ont pas lieu aussi souvent qu'ils devraient pour les patients hospitalisés à haut risque de mortalité. Une DOS est à la fois une responsabilité éthique et une intervention associée avec une meilleure qualité de soins en fin de vie. L'identification automatique des patients à haut risque de mortalité par un modèle de prédiction au moment d'admission pourrait stimuler des médecins à amorcer une DOS avec les patients identifiés.

L'objectif du projet présenté dans ce mémoire était de développer un modèle de prédiction de mortalité chez des patients hospitalisés qui soit automatisable et compatible avec les systèmes informatiques existants dans la région d'Estrie, précisément au Centre hospitalier universitaire de Sherbrooke (CHUS).

Nous avons appliqué l'expertise médicale et une technique d'apprentissage automatique supervisé, la forêt aléatoire, pour développer le modèle « Soutien à la discussion précoce à l'hôpital à l'aide du pronostic » (*Hospital Early Discussion Support Using Prognosis* ; HEADS-UP), en tenant compte des exigences d'un modèle de prédiction visé à être implanté dans un système d'aide à la décision clinique pour stimuler les DOS. Ces exigences incluent notamment celles liées à la validité externe, l'interprétabilité et l'utilité clinique d'un modèle de prédiction.

La validité externe et l'utilité clinique du modèle HEADS-UP étaient comparées à celles d'un modèle de référence, le modèle « Risque de mortalité hospitalière à un an modifié » (*Modified Hospital One-year Mortality Risk* ; mHOMR). Le modèle HEADS-UP avait une meilleure validité externe comparée au modèle mHOMR (indice C 0,89 vs 0,84). Le modèle HEADS-UP avait également plus d'utilité clinique que le modèle mHOMR en permettant d'identifier plus de patients qui n'ont pas eu de documentation minimale d'une DOS (la documentation du statut de réanimation) pendant une hospitalisation ni l'année précédant l'hospitalisation et qui sont décédés à l'intérieur d'un an de l'admission.

Ce projet contribue sur deux fronts au discours scientifique à propos de « l'intelligence artificielle » en médecine. Premièrement, il fait preuve des stratégies pour garder l'équilibre entre l'interprétabilité et la validité prédictive dans l'application de l'apprentissage automatique à des données cliniques hautement dimensionnelles, notamment avec une méthode de sélection de variables qui est dirigée par l'expertise médicale plutôt que l'algorithme d'apprentissage.

Deuxièmement, il démontre une méthode novatrice d'évaluation d'utilité clinique d'un modèle de prédiction automatisable qui vise à stimuler les DOS, en comparant la performance prédictive du modèle avec des issues axées sur le patient.

Remerciements

Merci à mes directeurs de maîtrise, Jean-François Ethier et Luc Lavoie, pour leur conseil à travers l'évolution de ce projet de recherche. Merci au personnel de l'InfoCentre du CHUS pour l'extraction diligente des données pour ce projet avec des ressources informatiques qui n'étaient pas conçues pour l'ordre de grandeur des données impliquées. Merci à Félix Camirand Lemyre et Martin Vallières pour l'intérêt qu'ils ont porté à l'évaluation de ce mémoire et leur aide pour l'améliorer.

Table des matières

Sommaire	iii
Remerciements.....	v
Liste des abréviations.....	ix
Liste des tableaux.....	xi
Liste des figures	xii
Introduction.....	1
Contexte	1
Objectif	2
Plan de mémoire	2
Chapitre 1 État des connaissances	4
1.1 Concepts fondamentaux	4
1.1.1 Cycle de vie.....	5
1.1.2 Propriétés centrales	7
1.2 Revue des modèles de prédiction de la mortalité pour informer une DOS.....	10
1.2.1 HOMR.....	11
1.2.2 Un modèle d'apprentissage profond pour soutenir les DOS	15
1.2.3 Autres modèles.....	17
Chapitre 2 Analyse des exigences.....	19
2.1 Exigences opérationnelles	19
2.1.1 Définition de cohorte	19
2.1.2 Modèle automatisé.....	19
2.1.3 Données accessibles au moment d'admission	20
2.2 Exigences fonctionnelles.....	20
2.2.1 Validité externe	20
2.2.2 Interprétabilité.....	20
2.2.3 Utilité clinique	22
2.3 Résumé des exigences.....	24
Chapitre 3 Méthodes	26
3.1 Sources de données	26
3.2 Cohorte.....	26

3.3	Méthode d'apprentissage.....	26
3.4	Prédicteurs.....	28
3.4.1	Diagnostic d'admission texte libre.....	30
3.4.2	Diagnostics antérieurs.....	31
3.4.3	Milieu de vie.....	33
3.4.4	Oxygène à domicile.....	33
3.4.5	Nouveaux prédicteurs.....	34
3.4.6	Données manquantes.....	34
3.4.7	Prédicteurs retenus au final.....	35
3.5	Modèle mHOMR comparatif.....	35
3.6	Environnement.....	35
3.7	Définition de l'échantillon d'entraînement et de validation.....	36
3.8	Paramètres de la méthode de forêt aléatoire.....	37
3.8.1	Échantillonnage des données longitudinales.....	37
3.8.2	Paramètre ntree.....	39
3.8.3	Paramètre mtry.....	41
3.9	Évaluation de la validité interne.....	43
3.10	Évaluation de la validité externe.....	43
3.11	Évaluation de l'impact des stratégies de sélection de variables.....	44
3.12	Évaluation de l'utilité clinique.....	45
3.12.1	Utilisation du statut de réanimation comme indicateur de processus de DOS.....	45
3.12.2	Définition des variables en regard des SR.....	46
3.12.3	Comparaison de l'utilité clinique.....	47
3.13	Analyses statistiques.....	47
3.13.1	Intervalles de confiance bootstrap.....	47
3.13.2	Test de permutation sur les différences.....	48
Chapitre 4 Résultats.....		49
4.1	Description de l'échantillon.....	49
4.2	Validation interne des modèles.....	50
4.3	Validation externe des modèles.....	51
4.4	Impact des stratégies de sélection de variables.....	58

4.5	Utilité clinique.....	59
Chapitre 5 Discussion		67
5.1	Validation temporelle.....	67
5.2	Validité locale vs faisabilité multicentrique.....	67
5.3	Communication d'information dans un SADC.....	69
5.4	Forces	70
5.4.1	Interprétabilité avec des données hautement dimensionnelles	71
5.4.2	Évaluation d'utilité clinique.....	73
5.5	Limitations	73
5.5.1	Absence des données paraclinique.....	73
5.5.2	Qualité des données disponibles lors de l'utilisation du modèle	74
5.5.3	Population captive.....	75
5.5.4	Indice de concordance.....	75
5.5.5	Portabilité.....	75
5.5.6	Validation interne minimale	76
Conclusion		78
Annexe 1 Diagnostic d'admission		79
Annexe 2 Diagnostics antérieurs		83
Annexe 3 Service d'admission.....		85
Annexe 4 Regroupement personnalisé.....		86
Annexe 5 Coefficients de régression du modèle mHOMR		88
Annexe 6 Importance des variables		90
Références.....		91

Liste des abréviations

AA	Apprentissage automatique
AVC	Accident vasculaire cérébrale
CCI	Charlson Comorbidity Index
CHSLD	Centre d'hébergement et de soins de longue durée
CHU	Centre hospitalier universitaire
CHUS	Centre hospitalier universitaire de Sherbrooke
CIM-10	Classification internationale des maladies, 10 ^e révision
CIRESSS	Centre informatisé de recherche évaluative en services et soins de santé
CIUSSS	Centre intégré universitaire de santé et de services sociaux
CLSC	Centres locaux de services communautaires
DCI	Dossier clinique informatisé
DOS	Discussion sur les objectifs de soins
FN	Faux négatif
FP	Faux positif
HEADS-UP	Hospital Early Discussion Support Using Prognosis
HOMR	Hospital One-year Mortality Risk
IC	Intervalle de confiance
IET	Intubation endotrachéale
MB	Mégabyte
MED-ÉCHO	Maintenance et exploitation des données pour l'étude de la clientèle hospitalière
MPOC	Maladie pulmonaire obstructive chronique
NPR	Ne pas réanimer
NPRI	Ne pas réanimer ou intuber
PARS3	Plateforme apprenante de santé et des services sociaux
RCR	Réanimation cardio-respiratoire
RNI	Ressource niveau intermédiaire
SADC	Système d'aide à la décision clinique
SIGDU	Système d'information de gestion des urgences

SNOMED-CT	Systematized Nomenclature of Medicine Clinical Terms
SQL	Structured Query Language
SR	Statut de Réanimation
TALN	Traitement automatique de la langue naturelle
VN	Vrai négatif
VP	Vrai positif
VPN	Valeur prédictive négative
VPP	Valeur prédictive positive

Liste des tableaux

Tableau 1. Les prédicteurs retenus pour l'entraînement du modèle	29
Tableau 2. Description de l'échantillon stratifié par l'issue de mortalité à un an	49
Tableau 3. Indicateurs de la validité interne de chaque modèle	51
Tableau 4. Indicateurs de la validité externe de chaque modèle.....	52
Tableau 5. Indicateurs de validité externe des modèles dans chaque service d'admission	55
Tableau 6. Importance des variables diagnostiques dans la validité du modèle HEADS-UP	58
Tableau 7. Validité du regroupement personnalisé comparé au Charlson Comorbidity Index	59
Tableau 8. Description de l'échantillon de validation stratifié par l'issue de statut de réanimation (SR) absent.....	60
Tableau 9. Description de l'échantillon de validation selon la dernière valeur du statut de réanimation	60
Tableau 10. Comparaison de la performance pronostique des modèles	61

Liste des figures

Figure 1. Répartition de la cohorte totale pour la validation interne et la validation externe des modèles	36
Figure 2. Histogramme de la fréquence de sélection d'un patient dans 500 échantillons bootstrap	38
Figure 3. Fréquence de sélection d'une visite d'un patient dans l'échantillon auquel un bootstrap est appliqué en fonction du nombre de visites du patient dans l'échantillon longitudinal	39
Figure 4. L'indice C du modèle forêt aléatoire en fonction du nombre d'arbres	40
Figure 5. Le score de Brier du modèle forêt aléatoire en fonction du nombre d'arbres	41
Figure 6. L'indice C du modèle forêt aléatoire en fonction du nombre de variables à considérer à chaque division d'un arbre.....	42
Figure 7. Le score de Brier du modèle forêt aléatoire en fonction du nombre de variables à considérer à chaque division d'un arbre	43
Figure 8. Pente de calibration des modèles utilisant la validation croisée avec l'échantillon d'entraînement	51
Figure 9. Histogramme du logit du risque prédit.....	52
Figure 10. Pente de calibration des modèles dans l'échantillon de validation	53
Figure 11. Pente de calibration des modèles dans l'échantillon de validation stratifié par service d'admission.....	54
Figure 12. L'indice C en fonction du mois.....	56
Figure 13. Le score de Brier en fonction du mois.....	57
Figure 14. Pente de calibration tracée pour chaque mois	58
Figure 15. Calibration du regroupement personnalisé comparé au Charlson Comorbidity Index	59
Figure 16. Nombre moyen d'alertes par jour en fonction du seuil décisionnelle	62
Figure 17. Valeur prédictive positive (VPP) des alertes pour la mortalité à un an en fonction du nombre moyen d'alertes par jour	63
Figure 18. Valeur prédictive positive (VPP) des alertes pour l'absence du SR en fonction du nombre moyen d'alertes par jour	64
Figure 19. Comparaison du bénéfice net des modèles avec une courbe décisionnelle.....	65

Figure 20. Dernière valeur du statut de réanimation (SR) documenté en fonction du risque de mortalité prédit de chaque modèle	66
---	----

Introduction

Contexte

Une discussion sur les objectifs de soins* (DOS) est un processus de communication ayant habituellement lieu dans le milieu hospitalier et qui s'instaure entre un patient (ou son mandataire) et ses fournisseurs de soins de santé; le but de la DOS est de s'assurer que les soins reçus par le patient sont en accord avec ses volontés [1]. Au Canada, les DOS n'ont pas lieu aussi souvent qu'elles le devraient chez les patients hospitalisés qui sont en fin de vie [2]. Chez cette population vulnérable au surtraitement, une DOS sert à la fois à assurer l'adéquation des soins intra-hospitaliers avec les préférences du patient et à informer le patient d'un mauvais pronostic dans le cas où il n'était pas déjà au courant. L'absence de cette discussion en temps opportun diminue la qualité des soins et peut résulter en des soins plus agressifs que ce qui est souhaité par le patient [3]. L'incertitude du pronostic telle que perçue par les cliniciens et le temps requis pour bien effectuer une telle discussion comptent parmi les facteurs qui limitent les DOS [4]–[7]. L'estimation du pronostic est évidemment nécessaire afin de prioriser la discussion des objectifs de soins. Néanmoins, les médecins ont tendance à être trop optimistes quant à la survie des patients [8], ce qui peut mener à une perte d'opportunités pour préparer les patients et leurs proches aux soins de fin de vie.

En parallèle à cela, des efforts régionaux, provinciaux et nationaux tentent d'améliorer l'accès aux services en soins palliatifs. En fin de vie, il est fréquent que la trajectoire de soins soit une escalade d'interventions agressives qui s'échelonne jusqu'au moment où le patient décède. Les cliniciens et patients reconnaissent que cette trajectoire de soins est discordante avec la préférence de la majorité de la population qui est de mourir dans un contexte palliatif où les soins visent avant tout le confort. Il faut donc, comme première étape, identifier les patients en fin de vie qui bénéficieront de soins palliatifs. Dans le contexte hospitalier, une DOS est l'étape qui devrait suivre immédiatement cette première étape d'identification. Il faut également statuer si le niveau de soins souhaités par un patient, lorsqu'il est mis au courant de son pronostic, concorde avec une trajectoire

* Ce processus est relié à la planification préalable des soins (PPS), mais ce dernier est censé avoir lieu dans le milieu communautaire et le but est d'anticiper un plan relatif à la situation où le patient ne serait plus apte à prendre des décisions par lui-même.

de soins palliatifs. De cette manière, l'identification des patients en fin de vie dans le milieu hospitalier pourrait faciliter deux tâches connexes : les DOS et les références appropriées en soins palliatifs.

Des modèles prédictifs sont couramment utilisés en médecine pour guider les décisions cliniques [9]–[11]. Un cas intéressant d'utilisation des modèles prédictifs est la prédiction du risque de mortalité afin d'identifier les patients en fin de vie pour mieux cibler les interventions palliatives. À cet effet, il existe plusieurs outils pronostiques [12]. Par contre, pour la majorité de ceux-ci les données doivent être entrées manuellement par le clinicien dans un outil papier ou un logiciel. Considérant que les cliniciens ont déjà un manque du temps pour les DOS, ajouter un processus non automatisé n'est certainement pas optimal. En contexte d'un système où les données d'un dossier clinique informatisé (DCI) étaient accessibles, un modèle pronostique qui y aurait accès pourrait automatiquement identifier les patients ayant un mauvais pronostic, puis un logiciel intégré pourrait signaler au médecin traitant qu'il faut considérer entamer une DOS. Un tel système d'aide à la décision clinique (SADC) a le potentiel d'améliorer la qualité des soins en fin de vie, car il peut augmenter la probabilité qu'un DOS a lieu et que les préférences des patients soient respectées. À notre connaissance, il n'existe pas à ce jour dans la région d'Estrie un SADC qui identifie les patients à haut risque de mortalité afin de stimuler une DOS. Un prérequis pour développer un tel système est la disponibilité d'un modèle de prédiction compatible avec les sources de données régionales.

Objectif

Notre objectif est de développer un modèle de prédiction de mortalité chez des patients hospitalisés qui soit automatisable et compatible avec les systèmes informatiques existants dans la région d'Estrie, précisément au Centre hospitalier universitaire de Sherbrooke (CHUS). Ce modèle doit satisfaire les exigences pré-implantation d'un SADC qui identifie les patients à haut risque de mortalité afin de stimuler une DOS.

Plan de mémoire

Le restant du présent mémoire est divisé en six parties : l'état des connaissances, l'analyse des exigences, les méthodes, les résultats, la discussion et la conclusion. Dans la première partie, une description sera faite de l'état des connaissances relatif au développement et à la validation des

modèles de prédiction cliniques. Ensuite, une description et une analyse seront faites des modèles de prédiction publiés dans la littérature médicale qui servent à identifier les patients hospitalisés à haut risque de mortalité et qui pourront servir comme point de départ pour un modèle appliqué dans le CHUS. Dans la deuxième partie, une analyse des exigences d'un modèle de prédiction dans le contexte d'un SADC sera présentée. Dans la troisième partie, les méthodes utilisées pour développer et valider un modèle satisfaisant lesdites exigences seront détaillées. Dans la quatrième partie, les résultats de validation seront présentés. La cinquième partie présentera la discussion quant à la pertinence de nos résultats de recherche et des travaux futurs qui sont envisagés seront énoncés par le fait même. La sixième et dernière partie du mémoire réitéra les résultats dans le contexte des thèmes soulevés dans l'introduction.

Chapitre 1

État des connaissances

1.1 Concepts fondamentaux

Un modèle de prédiction clinique est un modèle qui combine une ou plusieurs variables pour prédire des issues pertinentes des soins de santé d'un individu [13]. Le modèle en tant que tel est simplement une fonction qui associe une ou plusieurs variables d'entrée à une variable de sortie. Les entrées sont habituellement des facteurs de risque pour une issue donnée et la sortie est habituellement une estimation de la probabilité qu'un individu avec ces facteurs de risque expérimente cette issue [14]. En médecine, la majorité des modèles sont bâtis en utilisant l'apprentissage automatique supervisé (AA)[†] avec l'objectif de prédire d'issues cliniques discrètes (p. ex., la mortalité) [15]. Dans l'AA, il y a un ensemble d'observations qui est disponible pour lequel les valeurs d'entrée et de sortie sont connues; l'objectif est d'ajuster les paramètres du modèle pour prédire les sorties (c. à. d. les variables dépendantes[‡]) en utilisant les entrées (c. à. d. les variables indépendantes[§]) avec l'aide des méthodes statistiques traditionnelles (p. ex. régression linéaire) ou des méthodes d'apprentissage algorithmique (p. ex. arbre décisionnel) [16]. Cet exercice d'entraînement est fait afin que le modèle puisse fournir une prédiction lorsqu'il est généralisé à des observations où la valeur de la sortie est inconnue par l'utilisateur du modèle.

Après un survol du processus général qui mène à leur application, nous allons aborder trois propriétés centrales des modèles de prédiction : la validité, l'interprétabilité et l'utilité. Nous nous limitons pour ce mémoire aux modèles de prédiction qui font la classification.

[†] Nous allons considérer seulement l'apprentissage automatique supervisé, c'est-à-dire lorsque la bonne valeur de la réponse est connue *a priori*. Bien que l'apprentissage non supervisé puisse permettre de générer des hypothèses intéressantes, les modèles qui sont ainsi développés n'ont pas de rôle clinique démontré en ce moment.

[‡] Habituellement, il n'y a qu'une variable dépendante. Elle est souvent représentée par « Y » et elle est synonyme des termes « response », « label » et « outcome » dans la littérature scientifique sur les modèles de prédiction. Conceptuellement, Y est l'issue que l'on veut prédire, p. ex. le risque de mortalité.

[§] L'ensemble de ces variables, représenté mathématiquement comme un vecteur des variables, est souvent représenté par « X » et le nombre de composantes par la lettre « p ». Conceptuellement, chaque composante de X ($x_1, x_2 \dots x_p$) est une variable (p. ex. x_1 représente l'âge, x_2 représente le sexe, etc.), et on veut utiliser l'ensemble des p prédicteurs dans X pour prédire l'issue Y. Le vecteur X est synonyme avec le terme « feature vector » et une composante de X est synonyme avec les termes « feature » et « predictor » dans la littérature.

1.1.1 Cycle de vie

Le processus « du banc jusqu'au chevet » d'un modèle de prédiction clinique est composé de trois phases consécutives : le développement, la validation externe et l'évaluation de l'impact [13].

1.1.1.1 Développement

Il y a plusieurs actions qui doivent être faites dans la phase de développement : établir la tâche du modèle, disposer d'un échantillon d'entraînement approprié pour ladite tâche, définir la variable de sortie, définir les variables d'entrée et choisir la méthode d'apprentissage.

Un des choix les plus importants dans le développement d'un modèle de prédiction consiste à choisir quelle méthode d'apprentissage sera préconisée. Ce choix est en partie basé sur l'importance relative qu'on attribue à la validité et l'interprétabilité afin de maximiser l'utilité. Historiquement, la plupart des modèles étaient développés avec la régression logistique ou avec les modèles de risques proportionnels. De concert avec l'accès accru aux données dans les dossiers cliniques informatisés, il y a une augmentation en popularité des méthodes algorithmiques qui sont plus flexibles que les méthodes statistiques traditionnelles. Ces méthodes algorithmiques incluent notamment l'apprentissage profond et les forêts aléatoires. En enlevant des hypothèses *a priori* (p. ex. une hypothèse sur la forme de la fonction qui lie les variables indépendantes avec la variable dépendante), ces méthodes peuvent inclure plus de dimensions d'information avec moins de transformations et de sélections des variables qu'une méthode basée sur la régression. Un article précurseur qui porte sur les différences entre les méthodes d'apprentissage basées sur un modèle statistique et les méthodes d'apprentissage algorithmiques a été publié par Breiman en 2001 [17]. Une référence compréhensive des méthodes d'apprentissage automatique est donnée par Hastie et coll. [16]. À noter que chaque méthode d'apprentissage possède ses propres caractéristiques qui font qu'une certaine méthode est considérée être plus ou moins appropriée pour une tâche donnée. Un survol des considérations qui sont prises dans le choix de modèle pour des applications en médecine est donné par Rajkomar et coll. [15].

Par la suite, il faut entraîner le modèle avec l'échantillon d'entraînement et évaluer la validité interne du modèle avec un échantillon de test. La distinction entre l'échantillon d'entraînement et l'échantillon de test est souvent arbitraire; en général l'ensemble des observations disponibles pour le projet est séparé en deux avec un bloc d'environ 10-20 % qui sera utilisé exclusivement pour la validation interne. C'est à l'aide de ces observations qui ont été exclues de l'entraînement que la

vérification est faite afin d'estimer à quel point les prédictions du modèle sont extrapolables aux nouvelles observations. En pratique, cette phase de développement du modèle est itérative; l'évaluation du modèle à l'aide de l'échantillon de test peut influencer rétroactivement les choix initiaux. La validation qui est faite pendant cette phase fournit la *validité interne* du modèle, soit sa validité avec les données de développement. À noter qu'une bonne validité interne est un prérequis pour procéder à la validation externe, mais vu la nature itérative du processus et l'influence des résultats sur le développement lui-même, la validité interne doit être interprétée avec les réserves appropriées [18].

1.1.1.2 Validation externe

Après que le modèle final ait été développé, il est nécessaire de démontrer qu'il est valide dans le ou les milieux d'application. L'échantillon utilisé pour cette étape de validation (qui vérifie la *validité externe*) doit être suffisamment différent de celui utilisé dans validation interne puisque le but est de démontrer que le modèle est robuste aux différences entre le milieu de développement et le milieu d'application. Ce faisant, ce qui est *suffisant* dépend de cette différence entre le milieu de développement et le milieu d'application [19]. Dans le cas où il est prévu d'appliquer le modèle dans le même milieu, il est alors possible de le valider à l'aide d'un échantillon de test plus récent (*validation temporelle*). D'autre part, s'il est envisagé de généraliser le modèle à un autre milieu, il est nécessaire d'utiliser un échantillon de test provenant d'un autre milieu (*validation géographique*) [18]. À noter qu'une bonne validité externe du modèle est un prérequis à son application dans le milieu prévu. Par ailleurs, si ce n'est pas le cas, il faut retourner à la phase de développement soit pour ajuster le modèle avec les nouvelles données ou soit pour faire une révision plus complète [19]. Cette étape de validation externe est essentielle, car les opérations de développement peuvent mener à ce qu'on désigne comme « surajustement » (« overfitting »). Le modèle est alors très (trop) bien adapté au jeu de données de test, mais ce surajustement se fait au coup de son applicabilité dans un contexte plus large.

1.1.1.3 Évaluation de l'effet sur les soins cliniques

Pour justifier l'intégration d'un modèle à grande échelle dans les soins usuels, l'utilisation dudit modèle doit démontrer qu'il apporte un bénéfice, par exemple un bénéfice au point de vue des issues des patients. Une telle évaluation porte plutôt sur des paramètres du domaine clinique et ne

sera donc pas traitée en profondeur ici, mais le lecteur peut se référer à Steyerberg (2013) pour plus de détails [13].

Finalement, une fois qu'un modèle est intégré dans le système de santé, il est important de périodiquement révéifier sa validité [19]. Les traitements, les tests diagnostiques, l'organisation du système, les guides de pratique, etc. changent en continu. Il n'y a donc aucune garantie que l'association entre les prédicteurs et les réponses demeure stable. Ainsi, la vérification périodique permet d'identifier le moment où la validité du modèle diminue et qu'il est donc temps de mettre le modèle à jour.

1.1.2 Propriétés centrales

1.1.2.1 Validité

La *validité* d'un modèle de prédiction clinique réfère à la validité de ses prédictions, soit de déterminer à quel point les prédictions d'un modèle sont fiables et précises. Pour ce faire, lorsque le modèle estime la probabilité d'une issue donnée, cette validité est évaluée selon deux propriétés distinctes d'un modèle de prédiction : la discrimination du modèle de prédiction et la calibration du modèle de prédiction. Clarifions que ces propriétés de validités sont évaluées à la fois dans l'étape de validation interne et la validation externe, décrites précédemment. La validité interne estime la reproductibilité de ces propriétés dans l'échantillon de développement tandis que la validité externe réfère à la transportabilité de ces propriétés dans un échantillon différent [11, p. 329].

La discrimination d'un modèle de prédiction est sa capacité à différencier des observations avec l'issue (cas positifs) des observations sans l'issue (cas négatifs). L'indicateur qui est le plus souvent utilisé pour mesurer cette propriété est l'indice de concordance (l'indice C). Cet indicateur est intéressant du fait qu'il est indépendant de la prévalence des cas positifs, ce qui permet de faire la comparaison de la discrimination des modèles de prédiction qui ont été validés dans différentes populations.

La calibration d'un modèle de prédiction représente le niveau d'accord entre les prédictions du modèle et les issues observées. Calster et coll. ont proposé une hiérarchie de sous-propriétés par laquelle la propriété générale de calibration pourrait être divisée [20]; dans le contexte de prédiction de risque pour informer une décision clinique, ils argumentent qu'il est convenable de

mettre l'emphasis sur le type de calibration qu'ils nomment la calibration « modérée ». Selon leur hiérarchie, un modèle de prédiction démontre une calibration modérée si, parmi les patients avec le même risque prédit d'un événement, le taux d'évènement observé est égal à ce risque prédit. Par exemple, parmi 100 patients chez qui un modèle de prédiction de mortalité à un an prédit que le risque de cet évènement est de 10 %, le modèle démontre une calibration modérée si 10 patients décèdent à l'intérieur d'un an. Il est souhaitable, par contre, d'avoir un modèle qui démontre ce que les auteurs nomment la calibration « forte », qui peut être définie comme l'égalité entre le risque prédit et le taux d'évènement observé, mais cette fois pour chaque sous-groupe défini par les combinaisons uniques des valeurs des prédicteurs. Par exemple, si le modèle inclut l'âge (en chiffre entier) et le sexe comme prédicteur, cette égalité doit être observée pour chaque combinaison d'âge et de sexe. Les auteurs argumentent que d'évaluer la calibration « forte » n'est pas réaliste pour les types de données disponibles dans le milieu médical, où la variance entre individus est beaucoup plus importante que le nombre d'individus observés, menant à des sous-groupes trop petits pour l'estimation fiable. En attendant une technique d'estimation de calibration « forte » plutôt que « modérée » qui est applicable pour ces types de données, il est effectivement pratique de mettre l'emphasis sur la calibration modérée. À noter qu'il y a plusieurs manières d'évaluer la calibration modérée; une comparaison des méthodes est présentée par Steyerberg [11, p. 304]. La manière la plus simple et la plus efficace consiste à tracer une pente de calibration, qui permet de visualiser le taux d'évènement en fonction du risque prédit [11, p. 297]. Pour le restant du mémoire, nous allons utiliser le terme calibration pour référer spécifiquement à la propriété de calibration « modérée ».

Il existe des indicateurs qui combinent la calibration et la discrimination d'un modèle de prédiction, comme le score de Brier et le R^2 de Nagelkerke. Par ailleurs, ces indicateurs sont difficiles à interpréter dans le contexte de prédiction sur des individus [18].

1.1.2.2 Interprétabilité

Un des facteurs les plus importants à considérer pour un modèle de prédiction clinique est son *interprétabilité*. Le choix de méthodes disponibles en ce moment force souvent à faire un choix relatif entre maximiser l'interprétabilité ou la validité [21]. Plusieurs auteurs affirment que le besoin d'interprétabilité vient du fait que le modèle développé se base sur une formalisation incomplète du problème [22], [23]. Le but d'avoir un modèle qui soit *interprétable* est donc de

pouvoir vérifier si les exigences non formalisées dans le domaine d'application sont atteintes hors du contexte de développement. Bien qu'il semble simple à première vue, l'interprétabilité est un terme ambigu qui réfère, selon le contexte et les auteurs, à différentes propriétés d'un modèle de prédiction. Néanmoins, la réutilisation de ce terme demeure utile afin de faire le lien avec le corpus scientifique existant (qui l'utilise abondamment) et référer à un ensemble de propriétés similaires d'un modèle qui sont assemblées du fait qu'elles influencent la confiance des utilisateurs à prendre des décisions qui sont basées sur ce modèle [22]. Une discussion détaillée des propriétés qui tombe dans ces catégories est donnée par Lipton [22]. L'*interprétabilité* dans son ensemble est déterminée par la méthode d'apprentissage – certaines méthodes étant intrinsèquement plus transparentes et explicables que d'autres – mais aussi par le choix des prédicteurs, par l'objectif d'optimisation et par l'utilisateur du modèle [21], [22]. D'ailleurs, il convient de souligner que même un modèle de régression linéaire peut se révéler être *non interprétable* dans un contexte de décision clinique, si ce modèle utilise un grand nombre de prédicteurs, particulièrement lorsque des transformations complexes sont utilisées ou les prédicteurs sont corrélés.

1.1.2.3 Utilité

Une pratique fréquente dans l'évaluation des modèles de prédiction clinique consiste à se concentrer davantage sur la validité d'un modèle que sur son utilité. Néanmoins, bien qu'une bonne validité d'un modèle de prédiction soit nécessaire pour que ledit modèle soit utile, ce n'est pas suffisant. En effet, si un modèle de prédiction ne permet pas d'améliorer les issues cliniques, alors ce modèle de prédiction n'est pas utile dans le milieu clinique, cela peu importe ses mesures de discrimination et de calibration [24]. L'utilité est relative à une tâche donnée. Par exemple, un modèle peut être utile pour ajuster le risque dans une étude épidémiologique qui compare différentes populations, mais non utile pour améliorer les décisions prises par les cliniciens pour un programme de dépistage [24]. Lorsqu'un modèle de prédiction est censé influencer la prise d'une décision clinique, on définit l'utilité clinique comme la propriété du modèle de mieux classer les cas sur lesquels différentes décisions cliniques sont prises que le système sans le modèle [24]. « L'utilité » d'un modèle de prédiction dans ce mémoire réfèrera à « l'utilité clinique ». Il est évidemment souhaitable d'évaluer l'utilité d'un modèle de prédiction avant qu'il ne soit implanté à large échelle afin d'éviter d'investir des ressources substantielles pour opérationnaliser un modèle de prédiction inutile. L'utilité clinique d'un modèle est une fonction de sa validité (une combinaison de sa discrimination et sa calibration), du seuil décisionnel choisi

et du coût de chacune des conséquences possibles de chaque décision. Le seuil décisionnel est la probabilité d'une issue clinique (p. ex. la probabilité qu'un patient soit malade) au-dessus laquelle une action (p. ex. un traitement) est suggérée, car jugée pertinente. Typiquement, les cas de figure possibles sont les suivants : agir auprès d'un patient qui aura l'issue clinique – vrai positif (VP), agir auprès d'un patient qui n'aura pas l'issue clinique – faux positif (FP), ne pas agir auprès d'un patient qui aura l'issue clinique – faux négatif (FN), et ne pas agir chez un patient qui n'aura pas l'issue clinique –vrai négatif (VN). Dans cette évaluation, il est essentiel que le choix du seuil soit contextualisé en considérant le scénario médical dans lequel le modèle sera utilisé, ce qui nécessite souvent l'opinion des experts. S'il est faisable d'estimer le coût de chaque conséquence avec et sans le modèle, il serait théoriquement possible d'utiliser la théorie décisionnelle formelle. Cependant, cette approche est rarement pratique en médecine, car le coût d'une conséquence peut varier entre les patients et, ce faisant, il est difficile de faire la collecte d'un échantillon représentatif de tous les coûts [25], [26]. Une manière intéressante d'estimer l'utilité d'un modèle qui simplifie l'estimation du coût est le calcul du bénéfice net [27]. Un guide détaillé de cette méthode est donné par Van Calster et coll. [28]. Dans la section 2.2.3 (Utilité clinique), nous allons examiner les déterminants de l'utilité d'un modèle de prédiction clinique qui visent à stimuler une DOS.

1.2 Revue des modèles de prédiction de la mortalité pour informer une DOS

Le nombre de publications portant sur le développement et la validation des modèles de prédiction clinique dépasse grandement le nombre de publications traitant de l'impact de ces modèles [13], [29]. Ce faisant, il est donc nécessaire d'être très sélectif lorsque vient le temps de décider d'utiliser un modèle basé sur l'information qui est disponible dans la littérature. Dans le contexte du présent mémoire, l'objectif est de développer un modèle qui aura un impact clinique réel et non un modèle qui présente une bonne validité interne. Pour répondre à cet objectif, nous allons nous concentrer sur l'analyse de deux modèles de prédiction récemment publiés [30], [31]. Bien qu'ils aient le même objectif et la même population ciblée, le développement de ces deux modèles diffère à plusieurs niveaux, cela en plus d'avoir différentes limites. La fin de la présente section est l'occasion de décrire deux autres modèles qui sont répertoriés, mais qui ne sont pas applicables à notre objectif ni à notre population ciblée. Nous faisons le choix de présenter ces deux modèles,

car, bien que nous ne les utiliserons pas, il n'en demeure pas moins que ces modèles présentent des propriétés qui permettent d'alimenter notre réflexion.

1.2.1 HOMR

Le modèle « Risque de mortalité hospitalière à un an » (*Hospital One-year Mortality Risk* ; HOMR) est la première itération dans une famille de modèles similaires, la première itération ayant été initialement publiée par Carl van Walraven en 2014 [31]–[34]. Chaque version subséquente du modèle HOMR était modifiée afin de l'adapter à la disponibilité des données dans un nouveau milieu d'application. Parmi ces différentes versions, la dernière version du modèle HOMR, soit le modèle mHOMR, est la version qui apparaît être celle qui est la plus applicable pour notre cas d'étude; à notre connaissance, il s'agit du seul modèle de prédiction de mortalité opérationnalisé dans une SADC automatisée pour faciliter et soutenir une DOS au Canada. La première version de ce modèle (HOMR original) et la dernière version de ce modèle (mHOMR) sont présentées ci-dessous.

1.2.1.1 HOMR original

Le score de HOMR est un modèle de prédiction clinique qui utilise la régression logistique comme méthode d'apprentissage et qui est en partie conçu pour informer des décisions cliniques et en partie conçu comme outil de recherche pour l'étude des soins de santé [32]. Le modèle HOMR utilise des prédicteurs qui sont disponibles dans les banques de données administratives pour estimer la probabilité de mortalité des patients hospitalisés dans l'année suivant leur admission. En 2015, Van Walraven et coll. ont obtenu des résultats qui démontrent une meilleure validité externe du modèle HOMR dans plusieurs grandes cohortes lorsque comparées aux résultats qui étaient jusqu'alors publiés pour des modèles ayant un objectif similaire, soit celui d'estimer la probabilité de mortalité chez des patients hospitalisés. La statistique c de HOMR était de 0.89-0.92 selon la cohorte et la pente de calibration visuelle était raisonnablement bonne sauf pour ce qui est des probabilités prédites entre 70 % et 95 %. À noter que les auteurs ont utilisé la statistique de Hosmer-Lemeshow pour décrire la calibration; il s'agit d'une pratique courante, mais qui n'est toutefois plus recommandée [35].

La quantité élevée des variables nécessaires pour calculer le score de HOMR rend son utilisation manuelle irréaliste dans le contexte des soins médicaux usuels. Cependant, son intégration dans un DCI pour permettre un calcul automatisé du risque de mortalité pourrait avoir un impact positif

sur les DOS en permettant de faciliter l'identification des patients à haut risque de mortalité sans que le médecin ait à faire un effort supplémentaire à cet effet.

1.2.1.2 mHOMR

Récemment, une version modifiée de HOMR (mHOMR) a été faite pour évaluer la faisabilité d'une SADC pour stimuler une DOS chez des patients à haut risque de mortalité; le mHOMR n'utilise que des prédictors habituellement disponibles dans un DCI au moment de l'admission d'un patient et il a été développé puis intégré dans un SADC dans deux hôpitaux à Toronto [34]. Les investigateurs de mHOMR ont évalué la faisabilité de ce système à augmenter le nombre de discussions portant sur les objectifs de soins et des consultations en soins palliatifs chez des patients admis dans un service de médecine interne générale. À noter que la discrimination du modèle était similaire à celle du HOMR original. Les auteurs mentionnent que la calibration du modèle a été vérifiée, mais les résultats ne sont pas publiés. Le seuil décisionnel avait été déterminé avec l'aide de cliniciens faisant partie d'un groupe multidisciplinaire. Le choix de ce seuil reposait sur les deux facteurs principaux suivants : les ressources disponibles et le besoin d'éviter une désensibilisation aux alertes (*alert fatigue*) qui serait due à un nombre élevé de faux positifs. Utilisant un seuil de 0.21, le système avait une sensibilité de 59 %, une spécificité de 90 %, une valeur prédictive positive de 36 % et une valeur négative prédictive de 96 %. Lorsque le système identifiait un patient ayant un risque au-dessus du seuil, le médecin traitant dudit patient était avisé que son patient était identifié comme ayant un risque de mortalité élevé; le médecin traitant était avisé soit par l'entremise d'un message dans le DCI dans un centre ou soit par courriel dans l'autre centre. Le message suggérait aux médecins, si leur jugement clinique était en accord avec les résultats du modèle, d'adapter les soins actuels et la planification des futurs soins à un patient qui approche de sa fin de vie si ce n'était pas déjà le cas. Il est à noter que le message ne présentait pas le risque exact prédit par le modèle. Comme résultat, l'intervention était faisable avec un signal suggérant que les alertes automatisées intégrées dans le DCI peuvent modifier la pratique en augmentant le nombre de DOS et/ou des consultations en soins palliatifs, bien que l'étude ne fut pas conçue pour valider cette hypothèse.

Le modèle utilisé dans cette étude de *Wegier et coll.* est relativement simple et inclut les variables suivantes :

- l'âge du patient,

- le sexe du patient,
- le milieu de vie du patient,
- le service d'admission,
- si l'admission est une réadmission urgente 30 jours ou moins après une admission précédente,
- le nombre d'admissions par ambulance au cours de la dernière année,
- le nombre de visites à l'urgence au cours de la dernière année,
- le niveau d'urgence de l'admission,
- si l'admission est faite par ambulance et
- si l'admission est directement faite aux soins intensifs.

Le modèle mHOMR n'incluait pas les variables suivantes qui étaient prises en compte dans le modèle original, car ils n'étaient pas disponibles au moment de l'admission dans les DCI tel qu'utilisés dans les hôpitaux de l'étude : le diagnostic à l'admission, le besoin d'oxygène à domicile et l'Indice de comorbidité de Charlson (*Charlson Comorbidity Index* ; CCI). Aussi, puisque le modèle était seulement appliqué en médecine interne générale, la variable du service d'admission était fonctionnellement invariante.

Les travaux décrits dans le présent mémoire ont débuté avant que le modèle mHOMR ne soit publié. Ainsi, le modèle de référence était le modèle HOMR original. Tout comme il a été rapporté par les auteurs du mHOMR, nous avons constaté que l'intégration des données diagnostiques était problématique si le but était d'avoir un impact clinique sur les DOS. Néanmoins, le problème présentait aussi une autre dimension : l'inclusion de ces variables, lorsqu'elles sont disponibles, augmente considérablement le nombre de dimensions du modèle de prédiction, au-delà des tolérances usuelles d'un modèle de régression logistique, même avec un échantillon de quelques centaines de milliers d'observations.

Au CHUS, la disponibilité des données portant sur les comorbidités devienne disponible dans les bases de données administratives de trois à six mois après l'hospitalisation. Similairement, le diagnostic principal qui est indiqué sur la feuille sommaire – lequel réfère au diagnostic expliquant *a posteriori* l'hospitalisation – est disponible aussi dans un délai de trois à six mois. Il existe également un diagnostic d'admission qui est disponible au début de l'hospitalisation – lequel est le diagnostic expliquant *a priori* l'hospitalisation. Ce diagnostic inscrit par le médecin sur le

formulaire de demande d'admission est cependant saisi dans une variable en format de chaîne de texte libre et donc présentant plusieurs enjeux au niveau de son automatisation. Entre autres, la méthode de régression logistique, qui est la méthode utilisée par HOMR et mHOMR, n'est pas un bon choix pour modéliser le texte libre [36].

Pour permettre l'inclusion dans le modèle de l'information provenant d'autant de variables catégoriques (chaque diagnostic distinct étant sa propre dimension), les auteurs du modèle original ont « aplati » ces dernières dans des transformations « catégorique-à-linéaire ». Pour aplatir un sous-ensemble de variables catégoriques, les auteurs ont ajusté un modèle de régression logistique secondaire avec ces variables, puis ils ont inclus dans le modèle principal une fonction qui calcule le résultat linéaire (sur l'échelle des *log-odds*) du modèle secondaire. De cette manière, le modèle principal n'a qu'à ajuster un seul coefficient au lieu des nombreux coefficients qui étaient préajustés dans le modèle secondaire. Cet aplatissement constitue une heuristique qui est décrite dans la modélisation prédictive clinique et qui est généralement faite pour réduire le nombre de variables à un seuil tolérable au modèle de régression et à la taille de l'échantillon [18, Ch. 10]. Il est fortement probable, voire certain, qu'il existe une association entre le service d'admission, le diagnostic d'admission et les comorbidités d'un patient^{††}; cette association entre des variables indépendantes du modèle est cependant impossible à spécifier avec des termes interactifs lorsque ces variables sont « aplaties ». Une telle association sans spécification mène au problème de colinéarité (ou « multicollinéarité ») dans les modèles de régression logistique, qui diminue la reproductibilité du modèle dans des échantillons indépendants et rend complexe l'interprétation de ces coefficients [37]. D'autres options existent pour traiter ce type de problème. Les méthodes algorithmiques, comme la forêt aléatoire, présentent une plus grande tolérance aux problèmes à variable d'entrée présentant des dimensions élevées pour une taille d'échantillon donnée. Dans ce contexte, la technique d'aplatissement pour superficiellement condenser des dimensions dans un modèle de régression est moins nécessaire et ne résultera pas en une violation des bonnes pratiques présumées avec l'utilisation du modèle [17].

^{††} C. à d, avec la connaissance d'une valeur d'un de ces variables pour une hospitalisation donnée, il est possible de prédire les valeurs des autres variables avec une meilleure précision que par le hasard seul; p. ex., il y a de bonnes chances qu'un patient admis en pneumologie pour une exacerbation aiguë d'une maladie pulmonaire obstructive chronique (MPOC) est connu pour avoir la MPOC comme comorbidité.

En termes d'interprétabilité, la transformation extensive des prédicteurs avant leur inclusion diminue l'avantage d'un modèle relativement simple comme la régression logistique [22]. Dans le modèle HOMR original, plusieurs diagnostics à l'admission qui mènent à un mauvais pronostic vital – tel que la démence, l'AVC et l'exacerbation aiguë de maladie pulmonaire obstructive chronique – sont associés à un coefficient négatif, lequel devrait certainement faire hésiter un clinicien. Ce clinicien ne sera pas en mesure d'expliquer comment ces diagnostics réduisent le risque de mortalité, même si le processus par lequel ces coefficients étaient déterminés est entièrement transparent d'un point de vue algorithmique. Un autre exemple où la capacité d'interpréter le résultat est compromise dans le HOMR original est le regroupement des comorbidités basé sur un modèle de prédiction destiné à des fins administratives : le CCI. Cet indice regroupe des diagnostics ayant des pronostics respectifs qui sont souvent dissimilaires, par exemple : l'asthme avec la maladie pulmonaire obstructive chronique dans la variable « maladie respiratoire chronique », et le cancer thyroïdien avec le cancer pancréatique dans la variable « néoplasie non métastatique ». Ces paires regroupent des maladies qui sont associées à des espérances de vie très différentes. Le regroupement de comorbidités, même s'il ne fait pas sens au point de vue pronostic vital peut être acceptable pour l'ajustement du risque dans une étude épidémiologique. Au contraire, cette utilisation réduit l'acceptabilité du modèle lorsque ce dernier est appliqué au niveau individuel pour soutenir une décision clinique. La composition de variables comme « maladie respiratoire chronique » ou « néoplasie non métastatique » pose donc un problème d'interprétabilité et donc de confiance significative pour un tel modèle.

Avec leur modèle modifié, les auteurs de mHOMR ont contourné le problème de dimensionnalité élevé en excluant ces variables. Néanmoins, il est possible de supposer qu'un modèle qui inclurait ces variables pourrait présenter un avantage en termes de validité et d'utilité; les diagnostics d'un patient sont clairement des éléments influençant le pronostic vital et la nécessité de discuter des objectifs de soins. Le défi est donc de réussir à inclure ces variables d'une manière qui puisse maintenir un bon niveau d'interprétabilité.

1.2.2 Un modèle d'apprentissage profond pour soutenir les DOS

En 2017, *Avati et coll.* ont publié un modèle de prédiction qui utilise l'apprentissage profond pour estimer le risque de mortalité entre trois à 12 mois après une hospitalisation [30]. Le modèle incluait 13 654 dimensions d'information provenant des diagnostics, procédures, prescriptions et

rencontres dans l'année précédant une hospitalisation. Cependant, nous allons limiter notre analyse de cet article, car la quantité et les types de données informatisées nécessaires pour entraîner et utiliser le modèle dépassent ce qui est faisable pour la majorité des hôpitaux québécois [38]. Cela étant, nous présentons ci-dessous un commentaire succinct au sujet de l'évaluation du modèle pour informer l'état actuel d'apprentissage profond visant à soutenir les DOS.

Le modèle avait une bonne discrimination (AUC de 0,87 chez des patients admis), mais une calibration potentiellement problématique (p. ex. parmi les patients avec un risque prédit de 20 %, 10 % sont décédés). L'AUC était de 0,93 lorsque le modèle prédisait le risque pour des patients hospitalisés ainsi que non hospitalisés, mais cette statistique est difficilement interprétable, car, selon les auteurs, le modèle est censé n'être utilisé que pour des patients hospitalisés. Il convient de rappeler que mHOMR est un modèle plus simple de plusieurs ordres de grandeur et qui avait une discrimination de 0,89 chez des patients admis.

Les auteurs ont exclu des hospitalisations où le patient est décédé à l'intérieur d'une période de trois mois après son admission. Cette exclusion était faite pour l'échantillon de développement et de validation interne. Dans la vraie vie, il n'est toutefois pas possible d'appliquer un tel filtre au moment de l'admission. Ce faisant, leur évaluation du modèle est incomplète. Il est aussi possible de souligner des erreurs dans la conceptualisation qui a été faite. Par exemple, l'énoncé selon lequel les soins palliatifs sont plus bénéfiques s'ils débutent 3-12 mois avant le décès d'un patient est problématique. Premièrement, l'accès aux ressources en soins palliatifs nécessite dans plusieurs contextes une évaluation d'un pronostic vital de moins de 3 mois. Deuxièmement, ce n'est pas concordant avec l'objectif d'optimisation de leur modèle. En excluant les patients qui décèdent à l'intérieur d'une période de trois mois, le modèle n'est pas optimisé pour distinguer entre ces cas (décès à l'intérieur de trois mois) et les cas ciblés (décès entre 3-12 mois), lesquels peuvent vraisemblablement avoir des caractéristiques similaires.

Concernant l'interprétabilité, les auteurs comprennent que la nature « boîte-noire » d'un modèle d'apprentissage profond diminue la confiance des cliniciens à utiliser le modèle. Pour générer des explications *post-hoc*, les auteurs utilisent une technique d'explication locale : ils retirent séquentiellement chaque prédicteur d'un cas puis observent le changement sur la probabilité prédite. Les prédicteurs qui induisent le plus de changement sur le risque sont présentés à l'utilisateur. Malheureusement, en raison des prédicteurs utilisés par le modèle, cette approche ne

réussit pas à générer des explications utiles pour les cliniciens. Sept des 19 prédicteurs qui figurent parmi les 0,01% plus importants dans le calcul du pronostic d'un patient avec un risque de 90,9% sont des statistiques sommaires cryptiques, comme la variance autour du nombre de codes assignés par jour liés au type de rencontre, et l'intervalle du nombre de codes assignés par jour liés à une procédure. Un clinicien ne peut pas déduire une relation cause-effet entre ces variables et le pronostic alors il n'est pas plus capable d'accepter ni rejeter la conclusion du modèle basée sur ces variables. À minima, un gain de confiance est très improbable, mais en fait, une baisse de confiance est tout à fait possible.

Lorsqu'il est comparé à mHOMR, il apparaît que ce modèle nécessite plus de données et qu'il n'a pas démontré une meilleure validité, interprétabilité, ni utilité.

1.2.3 Autres modèles

Une brève analyse est présentée ci-dessous relativement à deux modèles de prédiction où la mortalité était évaluée, mais qui sont non-applicables pour notre cas d'utilisation.

1.2.3.1 Un modèle forêt aléatoire avec des données rétrospectives

Sahni et coll. ont développé un modèle de forêt aléatoire pour prédire la mortalité à un an utilisant l'information disponible à la fin d'une hospitalisation [39]. Dans le cadre de la présente recherche, le but est d'utiliser l'information disponible au moment de l'admission avec l'intention qu'un DOS ait lieu pendant que le patient est dans le milieu hospitalier où les ressources palliatives sont habituellement plus accessibles qu'en le milieu communautaire et surtout avec l'espoir de mieux adapter les soins de l'hospitalisation indexe selon les souhaits du patient. L'information disponible au début d'une hospitalisation est différente et généralement moins complète que l'information disponible à la fin. Il en ressort que la majorité de leurs prédicteurs ne sont pas applicables dans le cas étudié dans notre projet.

On remarque tout de même l'utilisation des forêts aléatoires comme méthode d'apprentissage et la discrimination élevée du modèle. Les propriétés des forêts aléatoires et leur convenance en regard des exigences de notre projet seront décrites dans le Chapitre 3 (Méthodes).

1.2.3.2 Un modèle d'apprentissage profond chez les patients avec une démence

Wang et coll. ont développé un modèle d'apprentissage profond pour prédire la mortalité à 6, 12 et 24 mois chez les patients avec une démence dans le but d'identifier des candidats pour des interventions palliatives précoces [40]. L'objectif du travail était surtout d'explorer comment le modèle d'apprentissage profond utilise les variables pour prédire la mortalité plutôt que de démontrer sa faisabilité sur le plan pratique. La restriction de la cohorte à des patients avec une démence n'était évidemment pas applicable au cas de la présente étude.

Il semble y avoir une contradiction importante quant à la validité publiée et l'utilité clinique de ce modèle. Ce modèle avait une excellente validité pour prédire la mortalité, mais la vérification du modèle par les auteurs révèle que les prédicteurs de loin les plus importants étaient la présence dans les notes cliniques des termes liés aux stades avancés des soins de fin de vie, comme « hospice » et « palliative »; le modèle semble donc identifier des patients pour qui ces enjeux ont déjà été reconnus et même documentés. Ce n'est donc pas utile pour stimuler une intervention palliative précoce. Cette contradiction souligne l'importance d'évaluer l'utilité clinique d'un modèle au-delà de sa validité s'il utilise la mortalité comme une issue intermédiaire pour le besoin d'une intervention palliative. Ceci est encore plus important lorsque certains prédicteurs du modèle sont directement liés à des soins palliatifs. Si ce n'est pas fait, la conséquence directe est qu'il est risqué de découvrir trop tard que l'impact clinique n'est pas aussi élevé que ce qui a été estimé.

Chapitre 2

Analyse des exigences

Cette section décrira les exigences de la solution que nous proposons dans le cadre de ce mémoire. Il convient de rappeler que l'objectif du présent mémoire est de développer un modèle de prédiction de mortalité chez des patients hospitalisés qui soit compatible avec les systèmes informatiques existants dans la région de l'Estrie. Ce modèle doit satisfaire les exigences d'un SADC qui identifie les patients à haut risque de mortalité afin de stimuler une DOS.

Les exigences qui sont décrites ci-dessous sont séparées en deux catégories : les exigences opérationnelles (ce qui est exigé par les contraintes opérationnelles) et les exigences fonctionnelles (ce qui est exigé par l'utilisateur). Chacune de ces catégories inclut différentes exigences. Les exigences opérationnelles sont : la définition de cohorte, l'automatisation du modèle et la disponibilité des données au moment d'admission. Les exigences fonctionnelles sont : un modèle avec une bonne validité externe, un modèle suffisamment interprétable et un modèle cliniquement utile.

2.1 Exigences opérationnelles

2.1.1 Définition de cohorte

La cohorte chez qui le modèle est applicable doit être bien définie. Nous allons utiliser la même définition de cohorte que HOMR afin d'être comparables : tout patient adulte hospitalisé dans un service non psychiatrique. L'exclusion des patients admis en psychiatrie dans le modèle HOMR était à l'origine motivée par l'accès plus restreint des données psychiatriques dans les bases de données administratives d'Ontario [32]. Même si l'accès n'était pas restreint, il en demeure qu'une DOS chez les patients admis en psychiatrie présente des enjeux éthiques, notamment la capacité moindre de juger les risques et les bénéfices des traitements futurs. Ainsi, l'exclusion des patients en psychiatrie dans notre cohorte est appropriée pour l'objectif clinique de stimuler les DOS.

2.1.2 Modèle automatisé

Afin d'avoir un système automatisé, toutes les variables utilisées par le modèle doivent être disponibles dans les bases de données électroniques qui sont accessibles dans le cadre des soins livrés par l'institution.

2.1.3 Données accessibles au moment d'admission

Le modèle de prédiction doit utiliser seulement des variables qui sont disponibles au moment de l'admission au CHUS. Le développement du modèle se faisant avec un jeu de données rétrospectives, il faut porter une attention particulière aux données utilisées pour l'entraînement et la validation. La disponibilité d'une variable donnée doit donc être simulée en utilisant le délai maximal entre une hospitalisation et sa disponibilité dans une base de données opérationnelle. Par exemple, puisqu'il peut s'écouler jusqu'à six mois entre la fin d'une hospitalisation et la saisie d'une comorbidité pour la rendre disponible dans la base de données MED-ÉCHO, la visibilité des comorbidités codifiées après une hospitalisation doit être décalée de six mois. Si un patient est réadmis à l'intérieur de ce délai, les données de comorbidités de l'admission précédente ne doivent pas être utilisées, car elles n'auraient pas été disponibles pour une utilisation en contexte réel (prospectif).

2.2 Exigences fonctionnelles

2.2.1 Validité externe

Le modèle doit démontrer une validité externe acceptable. Il n'y a pas de test objectif pour déterminer que la validité externe d'un modèle est acceptable, mais si les mesures de discrimination et de calibration sont raisonnablement similaires à HOMR et mHOMR, il est alors possible de conclure que la validité dudit modèle est acceptable. L'utilisateur doit être confiant que la fidélité du modèle dans l'échantillon rétrospectif soit généralisable aux cas actuels.

2.2.2 Interprétabilité

Le modèle doit être suffisamment interprétable. Tel que mentionné précédemment, il n'y a pas de consensus au sujet de la définition du terme *interprétabilité*, et cela encore moins en regard de sa formalisation. L'article de Lipton [22] identifie plusieurs *desiderata* souvent non formalisés d'un modèle, soit la confiance des utilisateurs, la transférabilité du modèle dans un milieu dynamique, l'informativité des prédictions, et la justesse des prédictions. Il convient de prendre en compte chacune de ces *desiderata* en fonction du cas d'utilisation qui est étudié dans le présent mémoire afin de définir l'exigence d'interprétabilité.

2.2.2.1 Confiance

Il est nécessaire que les utilisateurs (les cliniciens) aient confiance en ce modèle. Le cas d'utilisation dans le cadre du présent mémoire ne sera pas prescriptif; les cliniciens ont toujours recours à leur jugement clinique pour prendre une décision. Ce faisant, le modèle ne nécessite pas autant de confiance qu'il serait nécessaire d'avoir dans un cas d'utilisation où l'utilisateur devrait renoncer à son contrôle. Dans le cas qui fait l'objet de l'étude actuelle, il est raisonnable pour un clinicien d'avoir confiance en ce modèle si : l'exigence de validité externe est atteinte et si les erreurs du modèle sont détectables par un clinicien.

Cette dernière exigence implique que tous les prédicteurs doivent avoir un lien plausible avec le pronostic. Si le clinicien ne comprend pas comment un prédicteur pourrait être associé avec le devenir du patient, il ne sera pas capable d'identifier une erreur lorsqu'elle est due à ce prédicteur.

Il est possible qu'il y ait des erreurs qui soient présentes au sein des données administratives de telle manière qu'elles conduisent à des erreurs de prédiction pour certains patients, mais qu'elles ne soient pas si fréquentes qu'elles soient détectables dans la validation sommaire du modèle. Pour être capable d'identifier ce type d'erreur, il est nécessaire que les valeurs des prédicteurs qui sont utilisés par le modèle pour un cas donné soient accessibles au clinicien afin de lui permettre de détecter une différence entre les données utilisées par le modèle et les données les plus à jour recueillies par le clinicien. Cependant, cette exigence d'être capable de vérifier une valeur n'est pas aussi importante pour tous les prédicteurs. Bien que toutes les valeurs doivent être accessibles – en considérant que l'attention du clinicien et la quantité d'informations affichables sont limitées – il est préférable de prioriser leur visibilité. Ce faisant, plus une variable est susceptible d'être erronée ou importante dans le calcul du pronostic, plus il est préférable qu'elle soit visible d'emblée facilement au clinicien afin de maximiser la détection des erreurs importantes de prédiction. Il est donc désirable qu'il y ait une manière de classer des variables selon leur importance et que ce classement permette de prioriser l'affichage des valeurs importantes dans le calcul du pronostic vital.

2.2.2.2 Limitation des effets imprévus

La validité externe du modèle représente sa validité dans un environnement statique; il n'y a pas d'interaction entre le modèle et l'environnement. Lorsque le modèle sera transféré à un environnement dynamique, soit l'hôpital, il est nécessaire que les décisions qui sont basées sur le

modèle ne mènent pas à un changement d’environnement qui rend son utilisation plus nuisible que la pratique standard qui est en place sans le modèle. Dans le cadre du cas étudié dans le présent mémoire, il apparaît qu’une discussion portant sur les objectifs de soins ou une consultation en soins palliatifs, le cas échéant, ne peut pas être nuisible si cette discussion est jugée être appropriée par le clinicien. Cependant, il est impossible de prévoir toutes les interactions qui pourraient avoir lieu. Un système de surveillance qui compare les issues pertinentes des cas positifs et des cas négatifs permettra de vérifier cette exigence sur une base périodique. Un tel système est toutefois hors des exigences du projet de recherche actuel.

2.2.2.3 Quantité d’information

Comme le modèle mHOMR, le modèle proposé doit seulement communiquer si un cas est au-dessus d’un seuil de risque; il n’y aura pas d’exigence sur le modèle pour communiquer plus d’informations, par exemple, au sujet de la prédiction exacte.

2.2.2.4 Décisions justes

Le *désiderata* que les décisions prises automatiquement soient justes n’est pas applicable à un modèle non prescriptif. En ce sens, il n’y aura pas de décision clinique prise automatiquement.

2.2.3 Utilité clinique

Le modèle doit démontrer qu’il est cliniquement utile. Afin de comprendre cette exigence, il convient d’analyser les déterminants de l’utilité clinique d’une SADC qui génère des alertes pour stimuler des DOS. Comme discuté précédemment, l’utilité d’un modèle est une fonction de sa validité, du seuil décisionnel choisi et du coût de chacune des conséquences possibles de chaque décision. En supposant que les ressources nécessaires pour l’utilisation d’un modèle sont directement proportionnelles au nombre d’alertes générées, le seuil décisionnel (S) sera préétabli selon les ressources disponibles, comme avec mHOMR, en étant basé sur le nombre moyen d’alertes voulues par jour (N_{alerte}). L’Équation 1 démontre la relation entre ces variables :

$$\Pr(Risque\ Prédit \geq S) = \Pr(Alerte) = \frac{N_{alerte}}{N_{totale}} \quad (1)$$

Dans ce scénario, s’il existe un modèle de référence (M_r) déjà jugé utile, une manière d’évaluer l’utilité clinique d’un modèle candidat (M_c) est de comparer sa validité avec celle de M_r lorsque

N_{alerte} est égal, soit lorsque le seuil du M_r (S_{Mr}) génère le même nombre d'alertes (N_{alerte}) que le seuil de M_c (S_{Mc}). C'est-à-dire, lorsque l'Équation 2 est vraie :

$$\Pr(\text{Risque prédit} \geq S_{Mr} | M_r) = \Pr(\text{Risque prédit} \geq S_{Mc} | M_c) \quad (2)$$

Si M_c est strictement plus sensible et plus spécifique que M_r , il est probablement plus utile à ce niveau de ressources (car il y a plus de VP, plus de VN, moins de FN et moins de FP). D'autre part, si les deux paramètres sont moindres, il est alors probablement moins utile à ce niveau de ressources. Si un de ces paramètres est plus élevé et l'autre moins élevé, on ne peut pas directement comparer la validité et il faut prendre en compte le coût des conséquences possibles du système. Si on permet la simplification que le seuil décisionnel choisi soit lié au ratio des FP aux FN, cette dernière analyse pourrait alors être faite en comparant le bénéfice net avec une courbe décisionnelle [41].

Les modèles dans la littérature qui tentent de stimuler une DOS ou des interventions palliatives sont optimisés pour prédire la mortalité et non directement le besoin d'intervention (qui n'est pas un objectif réaliste avec l'état courant des données disponibles). Comme illustré par le modèle de Wang et coll., il faut être conscient que d'approximer l'utilité d'un modèle basé sur l'issue de mortalité pour stimuler une DOS peut mener à une utilité sensiblement plus basse que prévue. Au moment de la prise de décision, un patient « FN » ou « VP » n'est pas nécessairement un bon candidat pour initier une nouvelle une DOS et un patient « FP » ou « VN » peut être un candidat idéal. Par exemple, il pourrait y avoir :

- Un décès inattendu chez une jeune personne (un possible FN où l'importance d'un DOS n'est pas prévisible)
- Un décès attendu d'un patient déjà admis en soins palliatifs (un possible VP où un DOS a nécessairement eu lieu et n'est plus pertinent)
- Une survie de plus d'un an chez un patient avec une maladie dégénérative connue (un possible FP où une DOS est indiquée)
- Une survie de plus d'un an chez un patient avec un nouveau diagnostic de maladie terminale après l'admission (un possible VN où une DOS est indiquée).

Il est souhaitable d'avoir une idée de la capacité du modèle à améliorer les décisions cliniques avant d'investir des ressources dans une étude prospective, et surtout une implémentation à grande

échelle [42]. Une manière d’apprécier l’impact potentiel d’un modèle sur les décisions prises relatives aux DOS est en décrivant le taux historique de documentation des patients identifiés par le modèle dans un échantillon de validation externe. Par exemple, si presque tous les patients identifiés par le modèle ont historiquement eu une DOS documentée, il n’est probablement pas capable d’améliorer la décision d’avoir un DOS. Alternativement, si le taux de documentation est plus faible qu’attendu pour des patients identifiés par le modèle et réellement à haut risque de mortalité, il est raisonnable de croire que le modèle est capable d’améliorer la décision d’avoir un DOS. En d’autres termes, si le modèle est capable d’identifier les cas où une DOS devrait avoir lieu, mais historiquement n’avait pas eu lieu, il peut être considéré cliniquement utile. Cela dit, la vraie utilité du système qui emploi le modèle peut seulement être appréciée prospectivement en mesurant la différence dans les issues cliniques avec ou sans le système. Ces issues cliniques peuvent inclure le taux de documentation d’une DOS, la concordance des soins fournis avec les soins voulus selon la DOS et l’utilisation des ressources palliatives. La vérification de cet impact réel du système est hors des exigences du projet de recherche actuel.

Avec l’arrivée de mHOMR, il est nécessaire que le modèle proposé soit plus utile que ne l’est le modèle de référence; si ce n’est pas le cas, il n’y a pas de bonne raison d’utiliser le modèle proposé au lieu de mHOMR. La comparaison de validité empirique entre les deux modèles lorsqu’ils génèrent le même nombre d’alertes et la comparaison du bénéfice net avec une courbe décisionnelle peuvent être des méthodes de comparaison complémentaires dans le choix de modèle.

2.3 Résumé des exigences

Nos exigences pour un modèle de prédiction clinique qui prédit la mortalité à l’intérieur d’une période d’un an sont résumées dans la liste ci-dessous.

1. La prédiction de mortalité à un an doit être applicable à tout patient adulte hospitalisé dans un service non psychiatrique.
2. Les variables d’entrée doivent exister dans les bases de données électroniques accessibles à l’institution.
3. Les variables d’entrée doivent être disponibles à l’admission.
4. La validité externe du modèle doit être acceptable.
5. Les erreurs du modèle doivent être détectables par un clinicien.

6. L'association entre un prédicteur donné et la mortalité doit être plausible.
7. Les valeurs des prédicteurs utilisés par le modèle pour un cas donné doivent être accessibles.
8. Le modèle doit être cliniquement plus utile qu'un modèle de référence.

Une exigence souhaitable, mais non nécessaire est la suivante : le système doit être capable de classer l'importance des prédicteurs.

Chapitre 3

Méthodes

3.1 Sources de données

Toutes les données étaient extraites rétrospectivement avec les services de l'InfoCentre du CHUS [43], lequel utilise l'entrepôt de donnée « Centre informatisé de recherche évaluative en services et soins de santé » (CIRESSS) [44]. En temps réel, les prédicteurs proviendront de trois sources différentes qui sont accessibles au Centre intégré universitaire de santé et de services sociaux (CIUSSS) de l'Estrie - CHUS : la base de données du DCI utilisé au CHUS (Ariane), l'instance locale de la base de données administrative MED-ÉCHO [45], et l'instance locale de la base de données administrative de l'urgence SIGDU [46]. La variable prédite, soit la mortalité à l'intérieur d'une période d'un an, provient du registre provincial de décès auquel l'accès se fait par l'entremise de l'InfoCentre.

3.2 Cohorte

Toutes les variables brutes disponibles pour calculer le score de HOMR original ont été extraites pour toutes les hospitalisations de patients adultes dans un service non psychiatrique débutant entre le 1 janvier 2007 et le 3 mars 2019 au CHUS. Les identifiants de patient et de visite étaient anonymisés à la source. Puisque le nombre de visites à l'urgence et le nombre d'admissions par ambulance dans la dernière année nécessitaient une année complète d'observation, la première date d'admission retenue dans la cohorte était le 1^{er} janvier 2008 inclusivement. La dernière date de décès disponible dans le registre provincial pour cette cohorte était le 1^{er} juillet 2019, alors la dernière date d'admission retenue dans la cohorte était le 1^{er} juillet 2018 inclusivement.

3.3 Méthode d'apprentissage

Dans le cadre de la présente recherche, la forêt aléatoire a été choisie comme méthode d'apprentissage. Ce choix s'explique notamment par le fait que cette méthode permet d'inclure plus de dimensions d'information qu'avec régression logistique, cela sans avoir le risque d'invalider les hypothèses *a priori*. Aussi, cette méthode est moins susceptible au surajustement par rapport à l'apprentissage profond.

La théorie qui soutient les forêts aléatoires était rapportée initialement par Leo Breiman [47]. Une introduction qui relativise cette méthode comparée à d'autres méthodes d'apprentissage basé sur des arbres est donnée par James et coll. [48, p. 320], et une description plus en profondeur est donnée par Hastie et coll. [16, p. 587]. Nous présentons un survol de cette méthode et encourageons le lecteur intéressé à se référer dans un premier temps à un de ces deux publications selon leur niveau de familiarité en sciences mathématiques^{†‡}.

La forêt aléatoire est méthode d'apprentissage qui exploite la technique de « bagging » ; chaque arbre dans la forêt est un arbre de décision entraîné avec un échantillon *bootstrap* de l'ensemble d'entraînement total, et la prédiction pour un cas donné est obtenue en agrégeant les prédictions de tous les arbres. Chaque arbre est un sous-modèle avec une relativement forte *variance* et un relativement faible *biais* ; l'arbre apprend l'association entre les prédicteurs et l'issue prédite avec beaucoup de flexibilité (faible biais), mais est à risque d'apprendre des associations « bruyant », qui ne sont pas généralisables hors de son échantillon d'entraînement (forte variance). L'agrégation de plusieurs sous-modèles qui sont basés sur un échantillonnage aléatoire des cas (c. à d., « bagging » ; ***Bootstrap Aggregating***) permet de réduire la variance, en gardant stable le biais [16, p. 558], [48, p. 316]. La forêt aléatoire est caractérisée par une deuxième technique : les prédicteurs considérés à chaque nœud dans chaque arbre sont sous-échantillonnés aléatoirement parmi l'ensemble total des prédicteurs (pour des tâches de classification, le nombre de prédicteurs suggérés est la racine carrée du nombre de prédicteurs total). Cette perturbation dans la construction des arbres permet de les rendre plus indépendants les uns des autres, qui favorise davantage la réduction de variance lorsque les prédictions sont agrégées. Intuitivement, cette technique augmente « la diversité des opinions » des arbres, chacun avec une partie de la bonne réponse, et avec moins de chance que des prédicteurs importants pour la population domine sur des prédicteurs importants pour des sous-groupes.

D'un point de vue de la performance prédictive, l'aboutissement principal de la méthode de la forêt aléatoire est qu'elle n'est pas aussi flexible que d'autres méthodes, comme des modèles « gradient boosted » et l'apprentissage profond. La flexibilité d'apprentissage extrême qui est possible avec ces dernières méthodes les rend plus susceptibles au surajustement, mais ça leur permet

^{†‡} Ces livres sont disponibles gratuitement sur le site web d'un auteur : <https://web.stanford.edu/~hastie/pub.htm>.

d'apprendre des associations extrêmement complexes (p. ex., entre une image médicale et le diagnostic correspondant), limitées davantage par la puissance de calcul disponible (et en pratique, l'habilité des modélisateurs [16, p. 397]) que par la méthode d'apprentissage comme tel. Cela dit, la performance de la forêt aléatoire est souvent compétitive avec celui des méthodes plus flexibles pour une variété des tâches d'apprentissage [16, p. 590], incluant la prédiction de mortalité [49], sans autant d'effort nécessaire pour optimiser les paramètres d'apprentissage.

La méthode de la forêt aléatoire a quelques propriétés intéressantes d'un point de vue d'interprétabilité, lesquelles permettent de satisfaire les exigences mentionnées précédemment. Il est possible de recenser les cas historiques similaires à un cas index (le cas sur lequel une prédiction est générée en temps réel) en utilisant un indice de similarité (ou proximité) naturellement calculable avec la forêt aléatoire; pour deux cas, h et i , on peut définir leur proximité, une propriété symétrique, comme le nombre de fois que h et i sont dans le même nœud terminal dans la forêt, divisée par le nombre d'arbres afin de normaliser l'indice. Par ailleurs, une façon simple de classer l'importance des variables prédictes est en mesurant la taille d'erreur dans les cas « out-of-bag » (non échantillonné) pour chaque arbre lorsque les valeurs des prédictes sont aléatoirement permutées parmi ces cas; la différence dans le nombre de bonnes classifications dans un arbre lorsqu'on entre les données « out-of-bag » originales versus les données « out-of-bag » avec permutation de prédictes m , moyenné sur tous les arbres, constitue l'importance de prédictes m [16, p. 593]. Plus que la différence est grande, plus que cette variable était importante pour minimiser l'erreur de prédiction. La technique qui a été décrite à l'origine peut être biaisée en faveur des variables ayant plus de niveaux catégoriques, mais il existe des extensions à la technique originale qui permettent de réduire ce biais [50].

3.4 Prédictes

Comme point de départ au choix de prédictes, les variables du modèle HOMR original ont été examinées à l'aune de nos exigences. Les prédictes finalement retenus pour l'entraînement du modèle sont présentés dans le Tableau 1. Parmi les variables qui étaient incluses dans HOMR, celles qui n'ont nécessité aucune modification ou peu de modifications afin de satisfaire nos exigences sont les suivantes : l'âge au moment de l'admission, le sexe, le service responsable à l'admission, si l'admission était directement aux soins intensifs, le niveau d'urgence de l'admission, si l'admission était par ambulance, le nombre d'admission par ambulance dans la

dernière année, et le nombre de visites à l’urgence dans la dernière année. Vu la possibilité d’inclure plus de dimensions d’information que lors de l’utilisation d’une régression logistique, le nombre de niveaux catégoriques dans les variables « service d’admission » et « urgence d’admission » a été augmenté pour refléter une plus grande granularité.

Parmi les variables incluses dans le modèle HOMR, celles qui ont dû subir des modifications significatives afin de satisfaire nos exigences sont les suivants : le diagnostic à l’admission, les diagnostics sur la feuille sommaire, le besoin d’oxygène à domicile, et le milieu de vie. Ces variables sont des prédicteurs qui sont théoriquement associés avec la mortalité. Ce faisant, il était tout de même souhaitable de les inclure dans le modèle étudié dans le cadre de la présente recherche, mais avec certaines modifications, telles que décrites prochainement.

*Tableau 1. Les prédicteurs retenus pour l’entraînement du modèle. * Ensemble de variables*

Type	Variable	Précision
Entier	Âge	Âge au moment d'admission
Entier	Visites à l'urgence	Nombre de visites à l'urgence au cours de la dernière année
Entier	Admissions par ambulance	Nombre d'admissions par ambulance dans la dernière année
Entier	Semaines d'hospitalisations	Semaines complètes d'hospitalisation au courant des derniers 90 jours
Ensemble de booléens	Diagnostic d'admission*	147 variables booléennes (Annexe 1)
Ensemble de booléens	Diagnostics antérieurs*	84 variables booléennes (Annexe 2)
Catégorique	Sexe	M / F
Catégorique	Milieu de vie	Domicile / résidence / soins chroniques / inconnu
Catégorique	Urgence d'admission	Urgent / semi-urgent / élective / obstétrique
Catégorique	Service d'admission	26 niveaux (Annexe 3)
Booléen	Saison grippale	Si l'admission est entre le début de décembre et la fin de février
Booléen	Admission aux soins intensifs	Si l'admission actuelle est directement aux soins intensifs
Booléen	Réadmission urgente	Si l'admission actuelle est une réadmission urgente à l'intérieur de 30 jours
Booléen	Diagnostics antérieurs visibles	Si le patient a eu une hospitalisation qui s’est terminée il y a plus de six mois ou une visite à l’urgence qui a débuté il y a plus de deux semaines
Booléen	Admission par ambulance	Si l'admission actuelle est par ambulance

3.4.1 Diagnostic d'admission texte libre

La seule variable représentant le diagnostic d'admission disponible au moment de l'admission au CHUS est la transcription électronique par un commis du diagnostic écrit sur le papier d'admission par le médecin responsable. Lorsque le commis transcrit ce diagnostic dans le champ texte-libre du DCI, un menu de complétion automatique permet de choisir zéro à plusieurs codes CIM-9 préétablis ou d'inscrire une chaîne de caractères non standardisée. Cette liste ne semble pas avoir été mise à jour depuis plusieurs années et le choix de code, si un code est choisi, n'est pas validé par un archiviste. Il ne semblait pas non plus avoir de vérification de qualité sur le texte-libre car les fautes d'orthographe étaient prévalentes lorsqu'un code n'était pas choisi. Bien que ce diagnostic d'admission puisse être erroné ou même absent dans certains cas, nous avons supposé qu'il pourrait augmenter la validité globale du modèle; la plupart des diagnostics sont intelligibles et concordent avec le diagnostic principal validé sur la feuille sommaire à la suite de l'hospitalisation. Étant donné que la valeur de ce prédicteur sera visible par le clinicien, lequel saura le diagnostic réel, une erreur pourra être détectée par l'utilisateur relativement facilement. Le défi principal pour inclure cette information était de la représenter d'une manière appropriée pour la méthode d'apprentissage. En présence de fautes d'orthographe ainsi que de plusieurs abréviations à la pratique de la médecine, nous avons opté pour une méthode dirigée par la connaissance du domaine plutôt qu'une méthode naïve de traitement automatisé de la langue naturelle (TALN).

En utilisant la liste des diagnostics principaux les plus fréquemment présents sur la feuille sommaire, nous avons examiné les chaînes de caractères entières et les jetons individuels les plus fréquents dans les diagnostics d'admission. Dans un processus itératif dirigé par un expert du domaine, une liste d'entités fréquemment mentionnées dans le diagnostic d'admission a été élaborée : chaque diagnostic est mis en correspondance avec une expression régulière qui vise à être le motif nécessaire et suffisant pour identifier une référence à cette entité. Les types d'entités incluaient des diagnostics (p. ex. cellulite), des symptômes (p. ex. diarrhée), des signes cliniques (p. ex. tachycardie), des processus de maladie (p. ex. infection), des procédures (p. ex. remplacement articulaire), des manifestations de maladie (p. ex. tumeur) et des qualificatifs (p. ex. sévère). Lorsqu'il était approprié de le faire, une deuxième correspondance avec des codes

CIM-9 était faite pour les diagnostics et les symptômes. Bien que chacun de ces termes soit reconnaissable par un clinicien, ces derniers n'ont en soi pas nécessairement une connotation pronostique. Par contre, du point de vue du clinicien, le prédicteur « diagnostic d'admission » est composé de l'ensemble des termes relatifs à un patient donné, et cet ensemble est censé avoir un lien pronostic (p. ex. « sepsis » et « sévère »). La chaîne de caractères sources et l'ensemble de termes résultant étant visibles au clinicien, il sera capable de détecter la présence d'une erreur de correspondance. L'ensemble des variables finales dérivées du diagnostic d'admission texte libre seront référées par le terme « diagnostic d'admission ». La liste de ces variables et leurs définitions se trouve à l'Annexe 1.

3.4.2 Diagnostics antérieurs

Les diagnostics d'un patient du médecin traitant pendant une hospitalisation sont inscrits sur la feuille sommaire, laquelle est envoyée aux services des archives après que le congé d'hôpital ait été donné au patient. Ensuite, cela prend de trois à six mois pour que ces diagnostics soient validés, codifiés avec CIM-10 et rendus disponibles dans la base de données MED-ÉCHO par l'équipe des archives. Lorsqu'un patient est admis suite à un séjour à l'urgence, un diagnostic principal est codifié en CIM-10 par un agent administratif; cette codification est faite au courant des deux semaines suivant l'admission du patient. Un tel diagnostic codifié existe pour toutes les visites à l'urgence, mais il n'est pas disponible dans le même délai relativement court si le patient n'est pas par la suite admis à l'hôpital. Ce diagnostic est disponible dans la base de données SIGDU. Comme dans le cas des diagnostics qui sont inscrits sur la feuille sommaire, ce diagnostic n'est donc pas nécessairement disponible au moment de l'admission.

Malgré les délais mentionnés ci-dessus, nous avons supposé que les diagnostics qui sont codifiés rétrospectivement peuvent améliorer la validité globale du modèle en utilisant ceux qui proviennent d'une hospitalisation antérieure assez à distance pour avoir laissé le temps aux données d'être saisies. Cette information est cliniquement pertinente : lorsqu'un clinicien est consulté pour admettre un patient à l'hôpital, les feuilles sommaires sont souvent la source primaire des diagnostics sur laquelle le clinicien se base, cela même si la dernière visite du patient à l'hôpital remonte à il y a plus de six mois. Afin d'inclure l'information sur les diagnostics antérieurs dans le développement et la validation du modèle, la disponibilité de cette information était donc temporellement décalée pour simuler ce qui était visible à un système informatique au moment de

l'admission. Une période de recul de cinq ans était appliquée à la disponibilité des diagnostics pour limiter les diagnostics non pertinents, ce choix découlant de discussions avec un expert du domaine. La période de recul pour les diagnostics SIGDU était de six mois pour donner préséance au diagnostic principal sur la feuille sommaire. Après 6 mois, les diagnostics codifiés et révisés en regard de cet épisode de soin étaient disponibles à travers la feuille sommaire dans Med-Echo et présentaient une bien meilleure qualité. Au final, chaque hospitalisation était associée avec les diagnostics MED-ÉCHO des hospitalisations ayant débuté dans une période d'au plus cinq ans avant l'admission et s'étant terminée au minimum six mois avant l'admission. Les diagnostics SIGDU étaient similairement associés, cela en ce qui concerne les hospitalisations ayant débuté entre six mois et deux semaines avant l'admission.

Après l'application de la fonction de décalage des diagnostics, il y restait 5912 codes CIM-10 qui étaient présents parmi les diagnostics MED-ÉCHO et 642 codes CIM-10 qui étaient présents parmi les diagnostics SIGDU. Au lieu d'inclure chaque code distinct comme prédicteur, des groupes de diagnostics sémantiquement similaires et plausiblement liés au pronostic ont été formés. Ce regroupement était fait itérativement par un clinicien dont le but était de maximiser l'information disponible au modèle tout en minimisant le chevauchement entre les prédicteurs et en gardant les prédicteurs qui étaient plausiblement liés avec le pronostic. L'analyse de survie (modèle de Cox à risque proportionnel) de chaque code CIM-10 ajusté pour l'âge et le sexe était employée afin de visualiser les codes pertinents et d'assurer que les regroupements n'étaient pas trop dissimilaires au point de vue du pronostic. Les interactions entre les codes n'étaient pas vérifiées. Le regroupement prenait compte de la différence sémantique entre : un diagnostic principal sur la feuille sommaire d'il y a plus de six mois (l'ensemble « diagnostics principaux »), un diagnostic secondaire sur la feuille sommaire d'il y a plus de six mois (l'ensemble « diagnostics secondaires »), et un diagnostic d'admission à l'urgence depuis six mois (l'ensemble « diagnostics à l'urgence »). Par exemple, la mortalité à court terme est différente pour un patient avec une embolie pulmonaire documentée il y a trois ans sur la feuille sommaire et une embolie pulmonaire diagnostiquée à l'urgence il y a trois semaines. Bien que le code CIM-10 pour l'embolie pulmonaire est unique, les instances pour ces deux cas n'étaient pas mises en commun dans le même regroupement de diagnostics. Les prédicteurs finaux étaient une combinaison logique des diagnostics provenant de ces trois ensembles. La liste de ces prédicteurs, présentée avec les codes

CIM-10 de leur composition respective, se trouve à l'Annexe 2. L'ensemble des variables finales dérivées des diagnostics codifiés seront référées par le terme « diagnostics antérieurs ».

3.4.3 Milieu de vie

La provenance et la destination du patient, les variables visant à représenter le milieu de vie d'un patient avant et après une hospitalisation, sont saisies par des commis administratifs au moment de l'admission et du congé, respectivement. Cependant, il convient de souligner que malheureusement, la provenance est presque toujours la salle d'urgence (indiquée par « services d'urgences ») lorsque le patient n'est pas directement admis dans un lit d'hospitalisation. Il en résulte donc que la variable ne représente pas au final le milieu de vie du patient. Bien que la destination d'un patient à la suite d'une hospitalisation soit fiable, elle doit occasionnellement être corrigée par les archivistes par rapport aux informations entrées par le personnel hospitalier. La disponibilité de cette information corrigée requiert donc un délai similaire à celui pour les diagnostics sur la feuille sommaire (jusqu'à six mois). Comme en ce qui a trait aux diagnostics antérieurs, nous avons supposé que le milieu de vie d'un patient pourrait améliorer la validité du modèle même si l'information n'était disponible que pour certains patients. Lorsqu'un patient est relocalisé dans un milieu de vie offrant plus de support, il est rare qu'il retourne subséquemment dans un milieu requérant plus d'autonomie. Afin d'adapter cette variable, le milieu de vie était transformé en quatre niveaux: si une destination antérieure n'était pas visible (« inconnu »), si le patient avait un antécédent de soins de longue durée (« soins chroniques »), si le patient avait un antécédent de soins dans une résidence autre (« résidence »), et si le patient ne remplissait pas ces critères (« domicile »). La valeur « soins chroniques » signifie que le patient était congédié vers un centre d'hébergement de soins de longue durée (CHSLD) ou une ressource niveau intermédiaire (RNI) après une hospitalisation remontante à il y a plus de six mois. La valeur « résidence » signifie que le patient était congédié vers une résidence privée ou une résidence non institutionnelle après une hospitalisation remontante à il y a plus de six mois sans toutefois remplir le critère de « soins chroniques ». La valeur « domicile » signifie que la destination antérieure est visible et qu'il ne s'agit pas d'une institution incluse dans les autres niveaux.

3.4.4 Oxygène à domicile

Les données indiquant un besoin d'oxygène à domicile sont fractionnées à travers plusieurs instances de la base de données « I-CLSC » dans la région de l'Estrie. L'accès à ces données pour

tous les patients dans la région desservie par le CHUS nécessite la coordination entre plusieurs centres de services, une tâche démesurément difficile. Opportunément, le besoin d'oxygène à domicile est codifié sur la feuille sommaire avec le code « Z99.8 » par les archivistes. Nous avons inclus cette variable dans le modèle avec le même pipeline que les diagnostics antérieurs.

3.4.5 Nouveaux prédicteurs

Deux prédicteurs additionnels en lien avec le pronostique ont été inclus dans le modèle : le nombre de semaines complètes passées à l'hôpital au courant des derniers 90 jours, et si le moment d'admission était entre le début de décembre et la fin de février (cette période de l'année représente une approximation de la saison grippale). Ces prédicteurs étaient inclus, car ils avaient un lien plausible avec la mortalité (selon notre expérience clinique) ainsi qu'une association significative avec la mortalité dans une analyse univariée avec un modèle de régression logistique ($p < 0.001$).

3.4.6 Données manquantes

Nous avons opté pour l'inclusion des variables qui explicitent l'absence d'information au lieu d'imputer des valeurs qui peuvent tromper l'utilisateur. Il y avait trois variables où les données pouvaient être manquantes dans le système informatique, mais disponible au clinicien: les diagnostics antérieurs, le milieu de vie, et le diagnostic d'admission. Pour les deux premières variables, les données manquantes étaient explicitées avant leur inclusion dans le modèle, décrit ci-bas. Il n'y avait pas d'imputation faite pour le diagnostic d'admission.

Lorsque la variable binaire d'un groupe diagnostic donné est égale à *faux* (ou égale à 0, selon le type choisi), cela signifie que soit le patient n'a pas rempli les critères de ce groupe malgré la présence des codes diagnostiques visibles, soit qu'aucun code diagnostique n'est visible. Il est possible de connaître cette information au moment de l'admission; une variable binaire qui signifie la présence d'au moins un code diagnostique visible (soit MED-ÉCHO, soit SIGDU) était incluse dans le modèle (« diagnostics antérieurs visibles »). Si le patient a eu une hospitalisation qui s'est terminée il y a plus de six mois ou une visite à l'urgence qui a débuté il y a plus de deux semaines, la valeur de cette variable est égale à *vrai*, sinon elle est égale à *faux*. De cette manière, l'absence d'information sur une comorbidité est distinguée de la présence d'information qui infirme une comorbidité. Cette distinction est essentielle en contexte de raisonnement clinique.

Comme mentionné, lorsque le milieu de vie n'était pas disponible, la valeur de cette variable était « inconnu ». Une contrainte à l'utilisation de la valeur « résidence » était que la codification des destinations incluses dans ce groupe (résidence privée et résidence non institutionnelle) n'a débuté qu'en 2013. Avant cette classification, la destination de ces patients était notée comme « domicile ». Afin de s'assurer que les modèles étaient capables de reconnaître le niveau « résidence », tous les échantillons d'entraînement incluaient des observations avec cette valeur (c'est-à-dire que tous les échantillons d'entraînement incluaient des hospitalisations après 2013).

3.4.7 Prédicteurs retenus au final

Un total de 244 variables ont été incluses dans le modèle final. On y retrouve 84 variables provenant de l'information diagnostique codée (l'ensemble « diagnostics antérieurs ») et 147 variables provenant de l'information diagnostique non structurée à la saisie (l'ensemble « diagnostic d'admission »). Il y avait au moins une variable « diagnostics antérieurs » positive pour 40.8 % des visites, et au moins une variable « diagnostic d'admission » positive pour 89.8 % des visites. Il y avait au moins une variable soit « diagnostics antérieurs » ou « diagnostic d'admission » positive dans 93.8 % des visites.

3.5 Modèle mHOMR comparatif

Afin d'avoir un modèle de référence qui était potentiellement applicable au CHUS, le modèle mHOMR a été recalibré avec les données locales. La seule variable avec des données manquantes (qui n'était pas manquante dans l'échantillon de Wegier et coll.) était le milieu de vie d'un patient, lequel n'est pas nécessairement disponible au moment de l'admission au CHUS. Nous avons inclus la même variable pour le milieu de vie dans le modèle mHOMR que celui inclus dans la forêt aléatoire. De plus, le regroupement des services avait un niveau additionnel comparativement à mHOMR, soit le groupe « icu », et qui incluait les services des soins critiques et ceux d'anesthésie-réanimation.

3.6 Environnement

Le chargement des données était fait avec Python 3.6. Le nettoyage, transformation et persistance des données étaient faits avec PostgreSQL 10.2. L'analyse des données était faite avec R 3.6.2 dans l'environnement de développement intégré RStudio 1.2. La procédure de *bootstrap* à deux

étages (section 3.8.1) était implantée dans C++ et appliquée à partir d'environnement de R avec RCpp. La forêt aléatoire était entraînée avec la bibliothèque « ranger » [51].

3.7 Définition de l'échantillon d'entraînement et de validation

L'objectif de la validation externe était de déterminer la capacité de généraliser chaque modèle au CHUS dans le futur. Dans un premier temps, chaque modèle était entraîné avec les admissions entre le 1^{er} janvier 2008 et le 30 juin 2016 inclusivement (échantillon d'entraînement) puis chacun était validé avec les admissions entre le 1^{er} juillet 2017 et le 1^{er} juillet 2018 inclusivement (échantillon de validation). Cette partition temporelle est considérée supérieure à une partition aléatoire pour la validation d'un modèle de prédiction, car la transportabilité du modèle entre le milieu d'entraînement et le milieu de déploiement ne dépendra pas seulement des variations aléatoires (c. à. d., entre des échantillons tirés aléatoirement d'une population à un moment précis), mais à des variations non aléatoires comme les variations temporelles et les variations géographiques. [10], [52]. Une visite par patient au courant de la période d'entraînement était sélectionnée aléatoirement pour l'échantillon d'entraînement du modèle mHOMR et une visite par patient au courant de la période de validation était sélectionnée aléatoirement pour l'échantillon de validation. L'échantillon de validation était identique pour les deux modèles. Toutes les visites dans la période d'entraînement étaient incluses dans le modèle de forêt aléatoire. Une schématisation des liens entre ces échantillons est présentée dans la Figure 1.

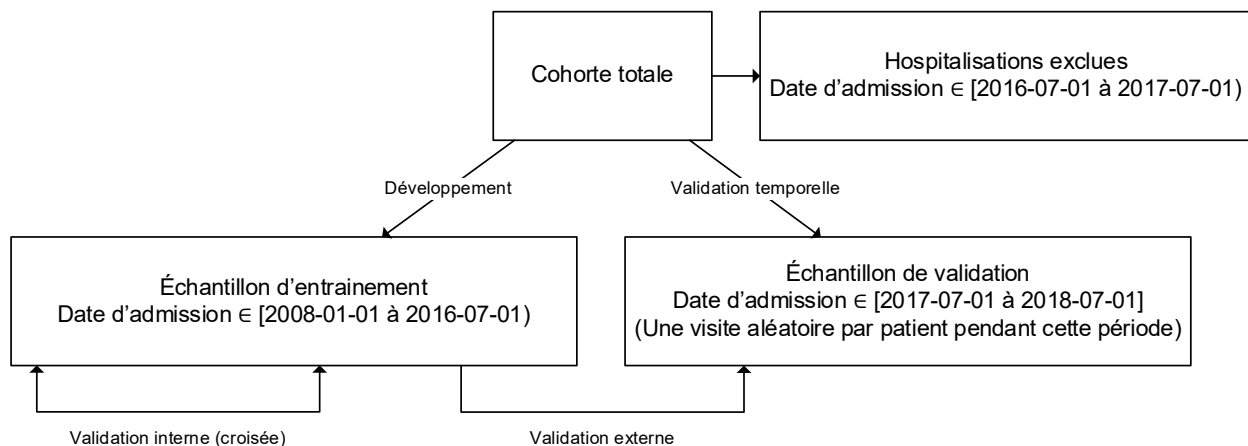


Figure 1. Répartition de la cohorte totale pour la validation interne et la validation externe des modèles

Les prochaines sections vont décrire les méthodes pour optimiser l'entraînement et évaluer la validité interne des modèles avec l'échantillon d'entraînement. Ensuite, les méthodes pour la validation externe avec l'échantillon de validation seront décrites.

3.8 Paramètres de la méthode de forêt aléatoire

La procédure d'échantillonnage des observations pour chaque arbre était modifiée pour prendre en compte le fait qu'un même patient pouvait être observé dans plusieurs visites (section 3.8.1). Le nombre d'arbres à utiliser (paramètre « *ntree* ») et le nombre de variables à considérer à chaque division d'un arbre (paramètre « *mtry* ») étaient le sujet des expériences, tel que décrit ci-dessous (sections 3.8.2 et 3.8.3). Le modèle final était une forêt aléatoire avec *ntree* fixé à 512 et *mtry* fixé à 15. Les autres paramètres étaient les valeurs par défaut pour la construction d'une « forêt de probabilité » [53]. Le nombre minimal d'observations dans chaque nœud terminal d'un arbre était 10 et la profondeur n'était pas limitée. Les détails se trouvent dans les sections suivantes.

3.8.1 Échantillonnage des données longitudinales

Puisque les données d'entraînement sont longitudinales, un patient pouvait être observé à plusieurs reprises. Plusieurs modèles dans la littérature, incluant les modèles HOMR et mHOMR, font la présélection d'une visite par patient dans l'échantillon d'entraînement, soit aléatoirement, soit en choisissant la première visite. Cette approche implique une perte importante d'information pertinente. Au lieu de faire une telle présélection de la forêt suivie par le *bootstrap* simple habituelle pour définir les cas « in-bag » de chaque arbre dans la forêt aléatoire, nous avons utilisé un échantillonnage *bootstrap* à deux étages [54] pour construire l'échantillon de chaque arbre directement à partir de l'échantillon longitudinal. Dans le *bootstrap* à deux étages, nous échantillonnons avec remplacement les patients distincts puis échantillonnons avec remplacement une visite pour chacun de ces patients. Cette modification permet de considérer toutes les visites d'un patient dans l'ensemble de la forêt tout en égalisant la probabilité qu'un patient donné apparaisse dans chaque arbre. La Figure 2 et la Figure 3 compare la fréquence de sélection des patients distincts et des visites distinctes, respectivement, lorsque nous faisons une présélection aléatoire d'une visite par patient suivie par un *bootstrap* simple comparé au *bootstrap* à deux étages. Le *bootstrap* à deux étages donne la même distribution de patients distincts, mais en incluant toutes les visites. On confirme que la probabilité de sélectionner une visite donnée d'un patient est inversement proportionnelle au nombre total de visites du patient (p. ex. si les visites

des patients avec une seule visite sont choisies en moyenne 500 fois, les visites des patients avec deux visites sont choisies en moyenne 250 fois chacune).

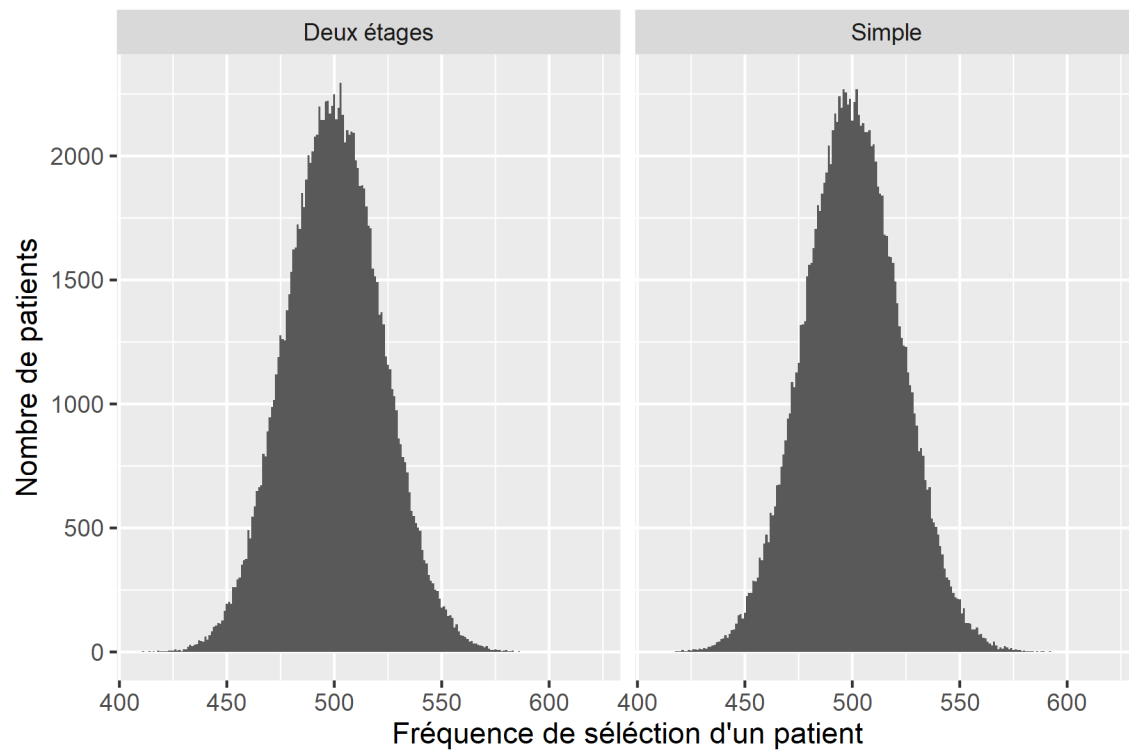


Figure 2. Histogramme de la fréquence de sélection d'un patient dans 500 échantillons bootstrap. Le bootstrap à deux étages appliqué à l'échantillon longitudinal est comparé à un bootstrap simple appliqué à un échantillon d'une visite par patient tirée aléatoirement de l'échantillon longitudinal.

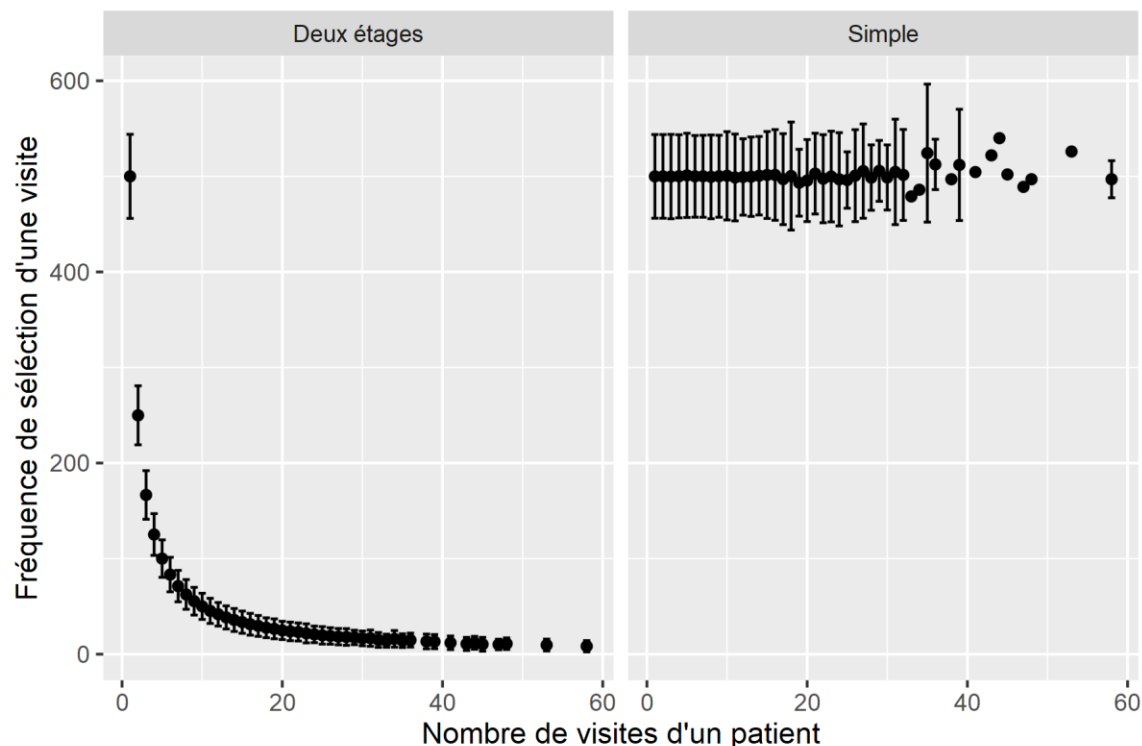


Figure 3. Fréquence de sélection d'une visite d'un patient dans l'échantillon auquel un bootstrap est appliqué en fonction du nombre de visites du patient dans l'échantillon longitudinal. Le bootstrap à deux étages appliqué à l'échantillon longitudinal est comparé à un bootstrap simple appliqué à un échantillon d'une visite par patient tirée aléatoirement de l'échantillon longitudinal.

3.8.2 Paramètre *ntree*

Pour trouver un nombre d'arbres (*ntree*) minimal, une forêt de 1024 arbres était entraînée sur les visites de 90 % des patients choisis aléatoirement dans l'échantillon d'entraînement. Le paramètre *mtry* était fixé avec une valeur par défaut de 15 (racine carrée du nombre total de prédicteurs). Des sous-ensembles de la forêt avec un nombre variable d'arbres compris entre 2 et 1024 étaient ensuite validés sur les visites des 10 % des patients restants. Pour chaque sous-ensemble, l'indice *C* (indicateur de discrimination) et le score de Brier (indicateur global de discrimination et calibration) étaient comparés (Figure 4 et Figure 5). Il avait peu d'avantages en regard de ces indicateurs à dépasser 128 arbres, mais aucune perte en validité n'était notée, cela jusqu'à 1024 arbres. En théorie, il n'y a pas de désavantage en termes de surajustement à utiliser plus d'arbres; nous considérons que 128 est la borne minimale pour notre cohorte, mais la borne maximale est plutôt une question de ressource. Finalement, une valeur de 512 arbres pour ce paramètre a été choisie en considérant le temps de calcul avec les ressources disponibles et la taille de la forêt résultante sur disque dur (l'entraînement d'une forêt aléatoire de 512 arbres avec 90 %

de l'échantillon prenait cinq minutes et générait un artéfact de ~500 MB; la construction des arbres était complètement parallélisée sur 16 noyaux de calcul à 3.7 GHz).

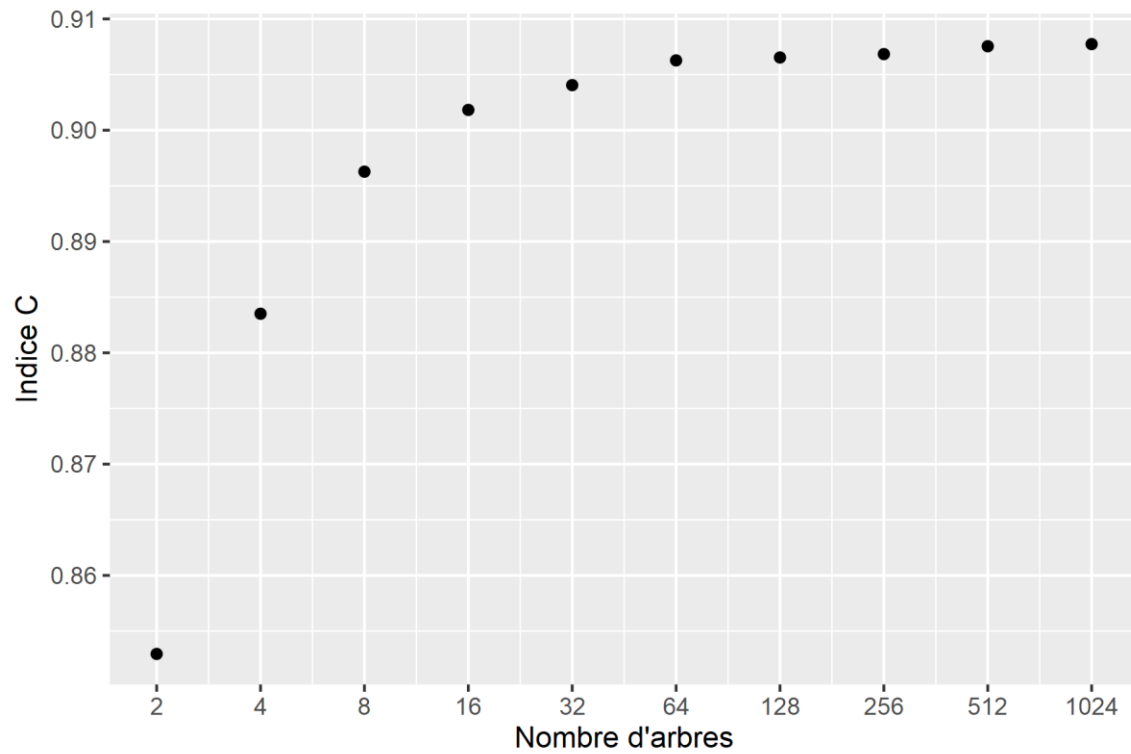


Figure 4. L'indice C du modèle forêt aléatoire en fonction du nombre d'arbres (paramètre *ntree*). Le nombre de variables à considérer à chaque division d'un arbre (paramètre *mtry*) est fixé à 15.

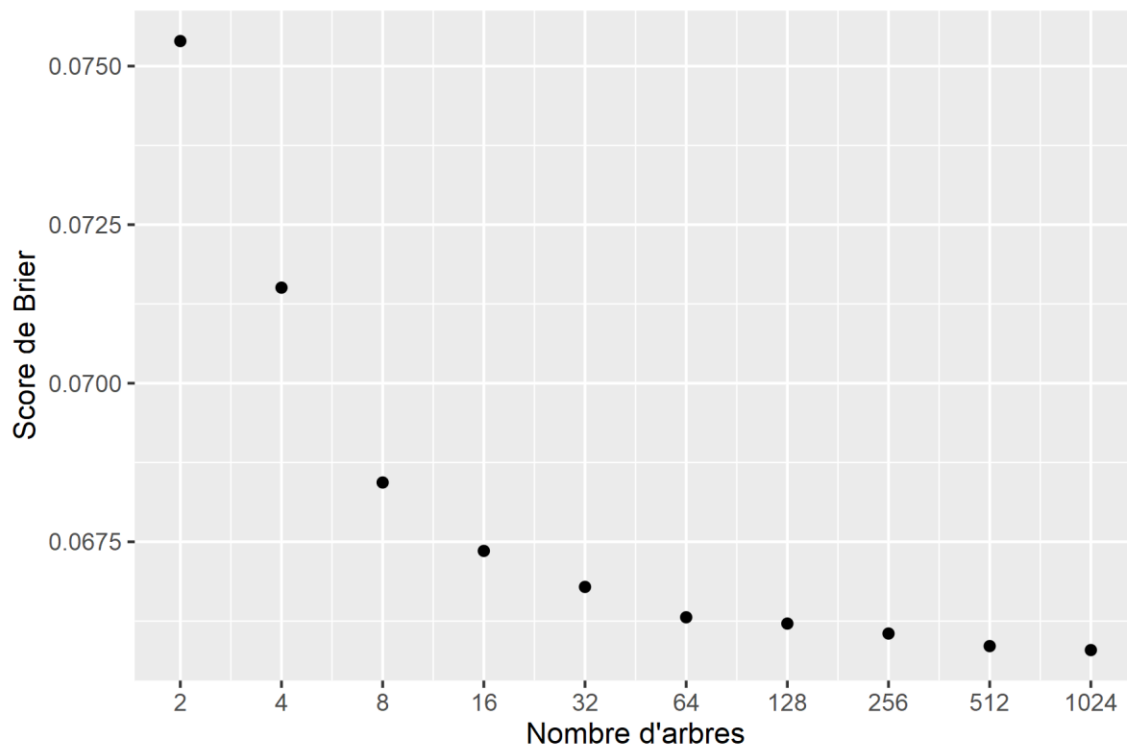


Figure 5. Le score de Brier du modèle forêt aléatoire en fonction du nombre d'arbres (paramètre *ntree*). Le nombre de variables à considérer à chaque division d'un arbre (paramètre *mtry*) est fixé à 15.

3.8.3 Paramètre *mtry*

Les auteurs de la méthode de forêt aléatoire ont suggéré que le nombre de variables à considérer à chaque division d'un arbre (*mtry*) soit par défaut la racine carrée du nombre de prédicteurs puis de l'ajuster au besoin [47]. Afin de s'assurer que cette valeur soit optimale, des forêts de 512 arbres étaient entraînées avec des valeurs de *mtry* comprises entre 2 et 64. Une recherche plus exhaustive autour de la valeur par défaut (15) a été effectuée. Les mêmes visites aléatoirement choisies parmi l'échantillon d'entraînement pour déterminer le paramètre de *ntree* ont été utilisées pour l'optimisation de *mtry*. Pour chaque forêt, l'indice C et le score de Brier étaient évalués (Figure 6 et Figure 7). Les valeurs de *mtry* autour de 15 étaient optimales pour les deux statistiques lorsque comparées aux deux extrêmes. En l'absence de variabilité importante pour les valeurs autour de la valeur par défaut, cette dernière (15) fut retenue.

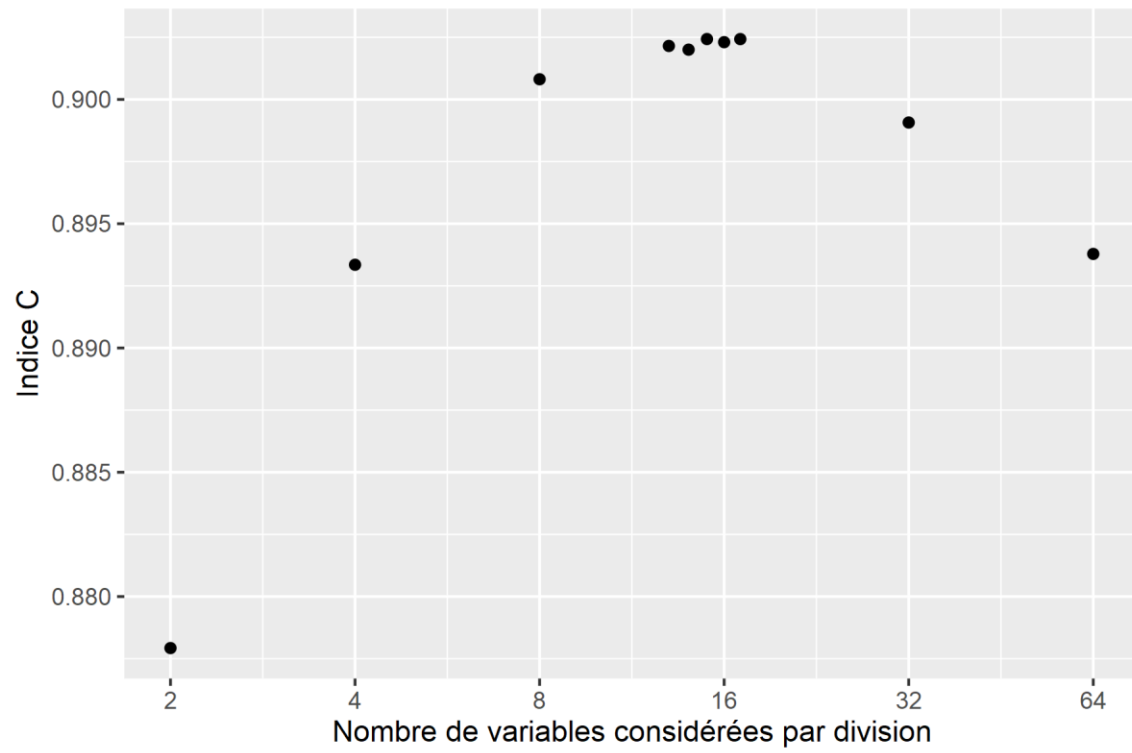


Figure 6. L'indice C du modèle forêt aléatoire en fonction du nombre de variables à considérer à chaque division d'un arbre (paramètre $mtry$). Le nombre d'arbres est fixé à 512.

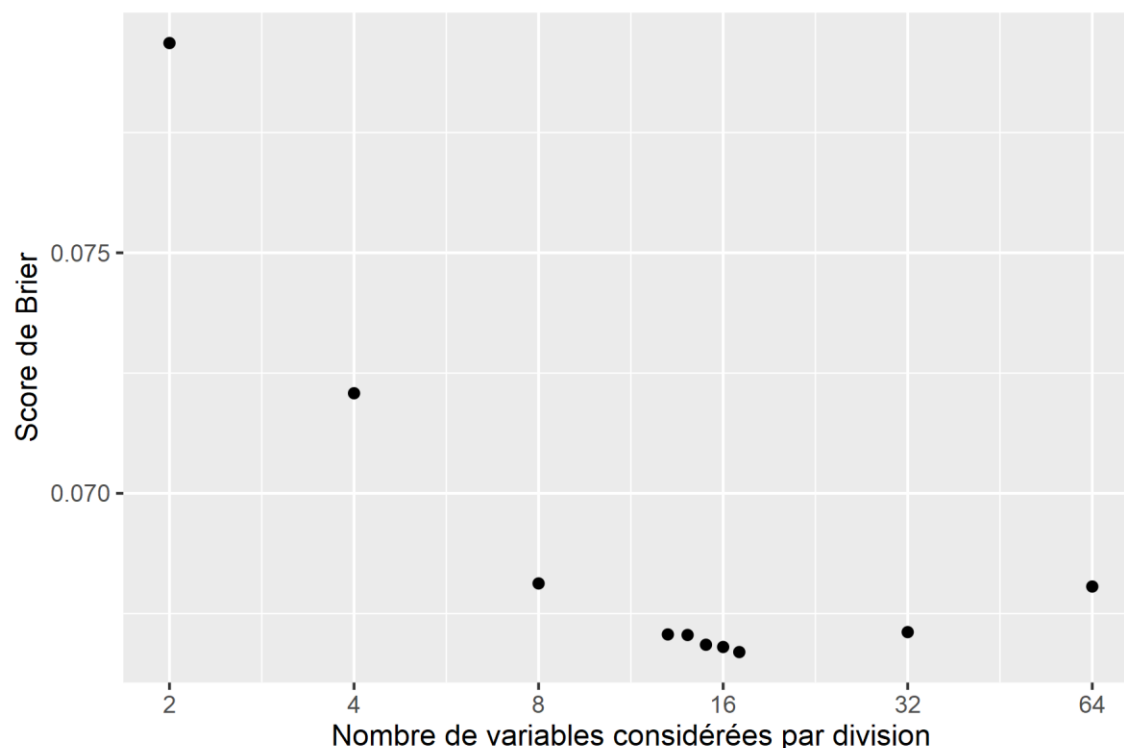


Figure 7. Le score de Brier du modèle forêt aléatoire en fonction du nombre de variables à considérer à chaque division d'un arbre (paramètre *mtry*). Le nombre d'arbres est fixé à 512.

3.9 Évaluation de la validité interne

La validité interne de la forêt aléatoire réalisée avec les paramètres mentionnés ci-dessus était évaluée avec une validation croisée à 10 plis. L'échantillon incluait toutes les admissions ayant eu lieu entre le 1^{er} janvier 2008 et le 30 juin 2016. Les patients distincts étaient divisés en dix groupes mutuellement exclusifs puis les visites de chaque patient étaient ensuite identifiées. De cette manière, aucun patient n'était présent dans plus d'un pli. L'échantillon de validation pour chaque pli contenait une visite aléatoire par patient. L'indice C, le score de Brier et la pente de calibration sont décrits dans la section des résultats.

3.10 Évaluation de la validité externe

L'indice C, le score de Brier et la pente de calibration sont rapportés pour chaque modèle afin d'évaluer leur validité temporelle, un type de validité externe, dans l'échantillon de validation. Ces indicateurs sont également décrits pour chaque service d'admission et pour chaque mois d'observation dans la période de validation (excluant juillet 2018, qui avait seulement une journée d'observation).

3.11 Évaluation de l'impact des stratégies de sélection de variables

Deux expériences ont été conduites pour évaluer l'impact de nos stratégies de sélection de variables en lien avec les diagnostics sur la validité du modèle.

Dans la première expérience, nous avons évalué la validité externe de la forêt aléatoire avec des sous-ensembles des prédicteurs finaux. Définissons les « prédicteurs de base » comme les prédicteurs finaux excluant ceux compris dans l'ensemble « diagnostic d'admission » et l'ensemble « diagnostics antérieurs ». Il y a eu trois modèles qui ont été entraînés : un modèle sans le diagnostic d'admission (prédicteurs de base + diagnostics antérieurs; 97 prédicteurs), un modèle sans les diagnostics antérieurs (prédicteurs de base + diagnostics d'admission; 160 prédicteurs), et un modèle sans le diagnostic d'admission et sans les diagnostics antérieurs (prédicteurs de base seulement : 13 prédicteurs). La validité de ces modèles partiels était comparée au modèle de forêt aléatoire final qui était évalué dans l'étape de validation externe.

Dans la deuxième expérience, nous avons comparé nos regroupements des comorbidités développés dans le cadre de l'entraînement de notre modèle avec ceux dans le CCI. L'expérience a été faite en tentant de simuler les conditions d'une étude épidémiologique rétrospective qui utilise les banques de données administratives où le délai de codification n'est pas pris en compte et les diagnostics sur la feuille sommaire à la fin d'une hospitalisation étaient associés à l'hospitalisation en question. La période de recul des diagnostics antérieurs était limitée à un an précédant l'admission. Les diagnostics à l'urgence étaient exclus (toutes les données provenaient de la base de données MED-ÉCHO afin de pouvoir évaluer le CCI, ce qui n'est pas possible avec un diagnostic unique d'admission à l'urgence). En regard des diagnostics provenant de MED-ÉCHO (donc saisis à partir des feuilles sommaires), il n'y avait pas de distinction faite entre le diagnostic principal et les autres diagnostics. Il n'y avait pas de combinaison logique des groupes provenant de différentes sources. Deux modèles qui incluaient l'âge, le sexe et les comorbidités étaient entraînés; la seule différence étant l'ensemble des groupes diagnostics utilisés (2 + 64 variables pour notre regroupement, 2 + 16 variables pour le CCI). Les groupes finaux sont dans l'Annexe 4.

Pour tous les modèles dans les deux expériences, l'échantillon d'entraînement et l'échantillon de validation étaient les mêmes que ceux dans la validation externe. Le paramètre *mtry* était le plancher de la racine carrée du nombre de prédicteurs. Tous les autres paramètres étaient identiques à la validation externe.

3.12 Évaluation de l'utilité clinique

Comme observé avec le modèle de Wang et coll., la validité d'un modèle servant à identifier des patients à haut risque de mortalité peut être trompeuse s'il est le seul critère pour juger l'utilité du modèle à stimuler un DOS. Le modèle mHOMR et le modèle proposé dans la présente étude ont un risque similaire à ce modèle d'avoir une validité qui surestime l'utilité. Par exemple, l'identification des patients dans le service des soins palliatifs augmente la valeur prédictive positive à tout seuil de prédiction, mais ne contribue aucunement à l'utilité du modèle; ces patients ont nécessairement eu une DOS. En ce qui concerne l'utilité clinique, nous estimons que l'évidence qu'un modèle identifie des patients qui ont un risque de mortalité élevé et une absence historique de DOS est supérieure à l'évidence qu'il identifie des patients à risque de mortalité élevée indépendamment de l'existence de DOS.

3.12.1 Utilisation du statut de réanimation comme indicateur de processus de DOS

Au CHUS, il n'est pas encore possible de documenter une DOS de manière électronique ni d'accéder automatiquement à la documentation des DOS ayant lieu hors de l'hôpital. Cependant, il est possible de documenter un statut de réanimation (SR) directement dans le DCI depuis 2015 pour toute visite à l'hôpital (incluant les visites ambulatoires). On remarque que l'absence d'un SR n'est pas automatiquement problématique à moins qu'elle soit accompagnée par un risque de mortalité élevé; chez des personnes qui sont jeunes et en bonne santé, une discussion du SR n'est pas habituelle ni requise (p. ex. chez une femme enceinte). Une telle discussion peut parfois générer une anxiété importante pour un bénéfice potentiel très limité.

Sachant que l'acceptation d'une réanimation cardio-respiratoire (RCR) implique l'acceptation d'une intubation endotrachéale (IET), car celle-ci fait partie de la RCR, la documentation électronique du SR comprend donc trois valeurs possibles : si le patient acceptera une réanimation cardio-respiratoire (RCR/IET), si le patient refusera une réanimation cardio-respiratoire, mais acceptera une intubation endotrachéale (NPR/IET), et si le patient refusera une réanimation cardio-respiratoire et une intubation endotrachéale (NPRI).

À noter que le SR est une partie de la DOS, mais une DOS comprend un périmètre beaucoup plus large. Si les objectifs de soins sont mis à jour dans le dossier médical papier, le SR doit être mis à

jour dans le DCI également. Bien que la politique du CHUS soit de documenter le SR à nouveau à chaque hospitalisation, une DOS est dans le cadre de la planification anticipée des soins, laquelle est un processus continu [1]; si un patient a eu une DOS pendant une visite, sa pertinence ne se termine pas subitement au moment du congé d'hôpital de ce patient. Dans un contexte communautaire, une mise à jour annuelle des objectifs de soins est recommandée chez des patients avec des maladies chroniques avancées [55]. Si la documentation d'un SR n'a pas eu lieu pendant une visite, et que ça n'a pas non plus eu lieu dans l'année précédant ladite visite, il est raisonnable de considérer qu'une partie des objectifs de soins applicables à l'état de santé du patient n'a pas été discutée, cela au moins d'une manière qui permet de vérifier la qualité des soins de manière durable.

Avec des données sur la documentation du SR en conjonction avec un modèle de prédiction, il est possible de répondre à la question suivante : est-ce que le modèle permet d'identifier des patients à haut risque de mortalité qui n'ont pas eu une documentation du SR? En posant cette question, il convient de rappeler que l'absence d'une documentation du SR est équivalente à l'absence d'une partie de la documentation des objectifs de soins. Il faut noter tout de même qu'il s'agit d'une approximation de la question clinique optimale qui serait en regard de la présence d'une DOS. La présence d'un choix en regard du SR n'implique pas qu'une DOS complète ait eu lieu. C'est au final la meilleure approximation clinique possible pour cette étude en regard des données disponibles.

3.12.2 Définition des variables en regard des SR

Tous les SR documentés électroniquement depuis leur initiation en février 2015, incluant pour des visites ambulatoires et des séjours à l'urgence, étaient extraits du DCI pour les patients dans notre cohorte. Deux variables étaient construites pour chaque hospitalisation : si le patient n'a pas eu un SR documenté pendant la période s'étendant entre un an précédant l'admission jusqu'au congé inclusivement (SR absent) et la dernière valeur documentée dans l'année précédente ou pendant la visite (Dernier SR : RCR/IET, NPR/IET, ou NPRI), cela si cette valeur était disponible. Cette dernière variable était construite pour décrire la préférence des patients selon le risque de mortalité prédit.

3.12.3 Comparaison de l'utilité clinique

Un SADC qui génère une alerte lorsque le risque de mortalité à un an est au-dessus d'un seuil décisionnel S était simulé avec chaque modèle pour les admissions entre le 1^{er} juillet 2017 et le 1^{er} juillet 2018 (échantillon de validation). Les modèles étaient identiques à ceux évalués dans l'étape de validation externe. L'utilité clinique était évaluée en comparant la validité des alertes par rapport à la mortalité à un an ainsi que la documentation du SR chez les patients identifiés par une alerte. Les modèles ont été comparés lorsque le nombre d'alertes qu'ils généraient était égal, c'est-à-dire lorsqu'ils supposaient la même disponibilité des ressources. Dans un premier temps, le seuil décisionnel S a été fixé à 0,21 pour mHOMR (tel que publié) et le seuil correspondant pour avoir le même nombre d'alertes était calculé puis fixé pour la forêt aléatoire. Les différences entre le VPP, le VPN, la sensibilité et la spécificité ont été analysées aux seuils respectifs de chacun des modèles. Le nombre moyen d'alertes par jour est présenté selon le seuil utilisé puis la proportion des cas avec l'issue de mortalité et absence du SR est présentée selon le nombre moyen d'alertes par jour. Le bénéfice net [27] lorsque le seuil décisionnel S varie entre 0,1 et 0,9 est également présenté. Finalement, la dernière valeur du SR lorsqu'il était documenté pendant une visite ou dans l'année précédant la visite est décrite en lien avec le risque prédit.

3.13 Analyses statistiques

3.13.1 Intervalles de confiance bootstrap

Pour les expériences de validation externe, lorsqu'une autre méthode n'est pas mentionnée, des intervalles de confiance de 95 % étaient générés avec 1000 répliques du *bootstrap*. Le but de la validation externe était d'évaluer la validité des modèles s'ils étaient entraînés avec toutes les données disponibles au CHUS avant minuit, le 31 juin 2017, puis appliqués prospectivement au CHUS pour une année. Ainsi, les modèles étaient entraînés une fois avec l'échantillon d'entraînement et appliqués une fois à l'échantillon de validation; chaque itération du bootstrap échantillonnée parmi les observations dans l'échantillon de validation et la statistique d'intérêt était recalculée. Les intervalles pour une statistique donnée étaient construits avec le 2,5^e et le 97,5^e percentile de sa distribution bootstrap dans l'échantillon de validation.

3.13.2 Test de permutation sur les différences

Un niveau de signification bilatérale à 0,01 était utilisé pour la comparaison de la validité des modèles de prédiction (6 indicateurs interreliés étaient comparés : l'indice C, le score de Brier, la VPP, la VPN, la sensibilité et la spécificité). Les tests statistiques sur des différences entre les modèles étaient faits avec un test de permutation standard [56] à 10 000 itérations. L'hypothèse nulle de ce test pour chaque indicateur était qu'il n'y avait pas de différence entre les modèles – entraînés avec les données théoriquement disponibles avant minuit, le 31 juin 2017 – dans l'échantillon de validation. L'hypothèse alternative était qu'il y'avait une différence. Pour chaque indicateur, la statistique de test était la différence dans cet indicateur calculé sur l'ensemble des prédictions d'un modèle versus l'autre. Si l'hypothèse nulle est vraie, pour chaque observation dans l'échantillon de validation, la prédiction émise par un modèle pour cette observation devrait être échangeable avec la prédiction de l'autre modèle pour cette observation. Ainsi, dans chaque itération du test de permutation, l'association entre une prédiction et le modèle qui a fait cette prédiction était aléatoirement permutée avant de recalculer la statistique de test. La valeur p était obtenue en comparant la statistique de test avec sa distribution dans les 10 000 échantillons de permutation, puis multiplié par deux pour le test d'hypothèse bilatérale.

Chapitre 4

Résultats

Pour le restant du mémoire, nous référons à notre solution comme le modèle « Soutien à la discussion précoce à l'hôpital à l'aide du pronostic » (*Hospital EARly Discussion Support Using Prognosis* ; HEADS-UP).

4.1 Description de l'échantillon

La cohorte incluse était composée de 125 082 sujets qui ont eu au total 260 940 hospitalisations entre le 1^{er} janvier 2008 et le 1^{er} juillet 2018 (Tableau 2). Dans 14,9 % des hospitalisations, le patient est décédé dans les 12 mois suivant l'admission (15,9 % des patients). L'âge moyen était de 60 ans (écart-type: 20) et les femmes représentent 54,6% des sujets. La moitié des visites étaient reliées à 20,8 % des sujets et 34,6 % parmi ces derniers sont décédés à l'intérieur d'une période d'un an après une visite. Ceci est tout à fait compatible avec un fait bien décrit, c'est-à-dire une augmentation marquée de l'utilisation de ressources de santé dans la dernière année de vie.

Tableau 2. Description de l'échantillon stratifié par l'issue de mortalité à un an. Les variables continues sont rapportées comme moyenne (écart-type). Les variables catégoriques sont rapportées comme fréquence (pourcentage).

Variable	Mortalité à un an	
	Faux	Vrai
Nombre de visites	222035	38905
Âge moyen (écart-type)	57.90 (20.37)	72.81 (14.30)
Sexe = Féminin (%)	124590 (56.1)	17953 (46.1)
Nombre moyen de visites à l'urgence (écart-type)	0.77 (1.48)	1.48 (1.98)
Nombre moyen d'admissions par ambulance (écart-type)	0.18 (0.64)	0.56 (1.20)
Saison grippale = Vrai (%)	55386 (24.9)	10057 (25.9)
Milieu de vie (%)		
Domicile	90130 (40.6)	21666 (55.7)
Résidence	2862 (1.3)	1557 (4.0)
Soins chroniques	1099 (0.5)	769 (2.0)
Inconnu	127944 (57.6)	14913 (38.3)
Nombre moyen de semaines d'hospitalisations (écart-type)	0.22 (0.94)	0.74 (1.54)
Urgence d'admission (%)		
Électif	47758 (21.5)	3819 (9.8)
Obstétrique	32605 (14.7)	10 (0.0)
Semi urgent	6707 (3.0)	779 (2.0)

Urgent	134965 (60.8)	34297 (88.2)
Admission par ambulance = Vrai (%)	56520 (25.5)	19211 (49.4)
Admission aux soins intensifs = Vrai (%)	3443 (1.6)	940 (2.4)
Réadmission urgent = Vrai (%)	16805 (7.6)	8147 (20.9)
Service d'admission (%)		
Cardiologie	29690 (13.4)	3405 (8.8)
Chirurgie cardiaque	3107 (1.4)	201 (0.5)
Chirurgie générale	20750 (9.3)	1406 (3.6)
Chirurgie maxillo-faciale	903 (0.4)	7 (0.0)
Chirurgie orthopédique	15585 (7.0)	1041 (2.7)
Chirurgie plastique	3721 (1.7)	82 (0.2)
Chirurgie thoracique	2553 (1.1)	301 (0.8)
Chirurgie vasculaire	7037 (3.2)	811 (2.1)
Endocrinologie	623 (0.3)	41 (0.1)
Gastro-entérologie	8048 (3.6)	2086 (5.4)
Gynécologie	8282 (3.7)	709 (1.8)
Hématologie / Oncologie	4536 (2.0)	3262 (8.4)
Médecine familiale	19329 (8.7)	6096 (15.7)
Médecine interne	19576 (8.8)	5630 (14.5)
Néphrologie	2982 (1.3)	1004 (2.6)
Neurochirurgie	7343 (3.3)	3454 (8.9)
Neurologie	6118 (2.8)	1218 (3.1)
Obstétrique	34420 (15.5)	16 (0.0)
Ophthalmologie	1624 (0.7)	49 (0.1)
Oto-rhino-laryngologie	4655 (2.1)	387 (1.0)
Pneumologie	7355 (3.3)	4663 (12.0)
Rhumatologie	907 (0.4)	98 (0.3)
Soins intensifs	1911 (0.9)	972 (2.5)
Soins palliatifs	46 (0.0)	969 (2.5)
Trauma	1488 (0.7)	146 (0.4)
Urologie	9446 (4.3)	851 (2.2)
Diagnostics antérieurs visibles = Vrai (%)	99558 (44.8)	25730 (66.1)

4.2 Validation interne des modèles

Les indicateurs de la validité interne du modèle HEADS-UP et du modèle mHOMR dans l'échantillon d'entraînement sont présentés dans le Tableau 3 et leurs pentes de calibration dans la Figure 8.

Tableau 3. Indicateurs de la validité interne de chaque modèle. La validité apparente était calculée en entraînant un modèle sur l'échantillon d'entraînement puis en appliquant le modèle sur ce même échantillon d'entraînement après l'échantillonnage d'une visite aléatoire par patient. La validité interne était calculée avec une validation croisée à dix plis. Les statistiques étaient calculées pour chaque pli après l'échantillonnage d'une visite aléatoire par patient. Ils sont rapportés sous forme de médiane (minimum, maximum).

	HEADS-UP	mHOMR
Validité apparente		
Indice C	0.946	0.860
Score de Brier	0.0552	0.0805
Validité interne (intervalle)		
Indice C	0.906 (0.901-0.914)	0.859 (0.852-0.873)
Score de Brier	0.0671 (0.0654-0.0697)	0.0803 (0.0784-0.0836)

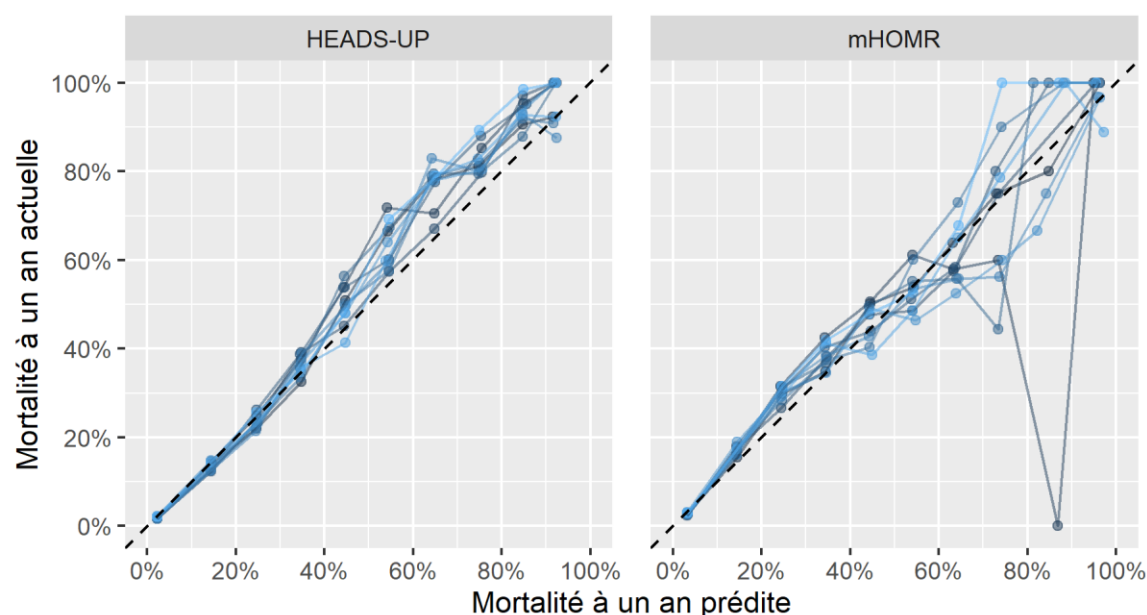


Figure 8. Pente de calibration des modèles utilisant la validation croisée avec l'échantillon d'entraînement. La proportion des cas réellement décédés à l'intérieur d'un an est présentée pour chaque décile de risque prédit et pour chaque plis. La ligne pointillée représente une calibration idéale.

4.3 Validation externe des modèles

Chaque modèle était entraîné avec les données disponibles à l'admission pour 105 275 patients hospitalisés entre le 1^{er} janvier 2008 et le 1^{er} juillet 2016. Le modèle HEADS-UP échantillonnait parmi les 208 933 visites de ces patients tandis qu'une visite par patient était choisie aléatoirement pour entraîner le modèle mHOMR. Les coefficients du modèle mHOMR ajustés avec les données locales étaient majoritairement dans la même direction que ceux ajustés avec les données

ontariennes (Annexe 5). La cohorte de validation comprenait les données disponibles à l'admission pour 19 992 patients hospitalisés entre le 1^{er} juillet 2017 et le 1^{er} juillet 2018 (une visite aléatoire par patient). La distribution du *logit* du risque prédit est présentée à la Figure 9.

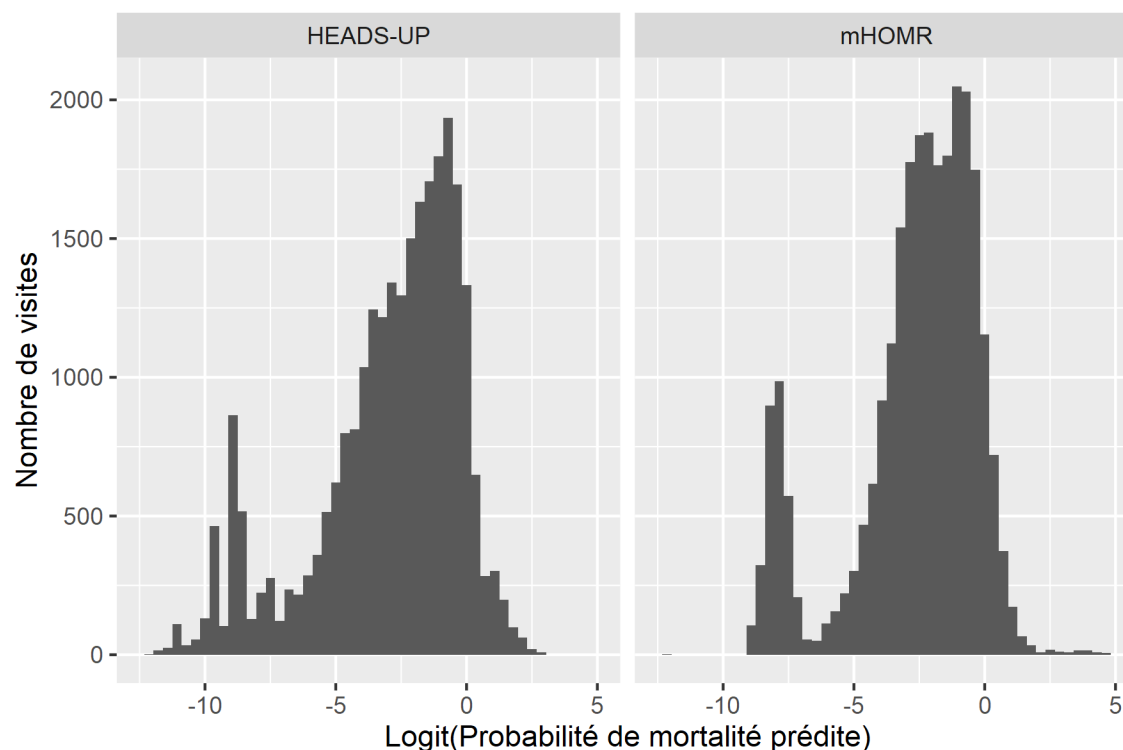


Figure 9. Histogramme du logit du risque prédit. $\text{Logit}(p) = \ln(p/(1-p))$.

L'indice C du modèle HEADS-UP était 0.888 (IC 95 % : 0.882 – 0.894) et celui du modèle mHOMR était 0.838 (IC 95 % : 0.831 – 0.845) (Tableau 4). Il est à noter que l'indice C et le score de Brier du modèle HEADS-UP étaient légèrement inférieures lorsque comparés à la validation interne, mais les deux étaient supérieures lorsque comparés au modèle mHOMR. Les deux indicateurs étaient supérieurs pour le modèle HEADS-UP dans tous les services à part deux exceptions, les soins palliatifs et le trauma, qui avait relativement peu d'observations (Tableau 5). Les pentes de calibration globale pour chaque modèle sont présentées dans Figure 10 et celles stratifiées pour le regroupement des services du modèle mHOMR sont présentées dans Figure 11.

Tableau 4. Indicateurs de la validité externe de chaque modèle

	HEADS-UP	mHOMR	
Indice C (IC 95%)	0.888 (0.882 - 0.894)	0.838 (0.831 - 0.845)	p < 0.001
Score de Brier (IC 95%)	0.0744 (0.0722 - 0.0767)	0.0860 (0.0835 - 0.0885)	p < 0.001

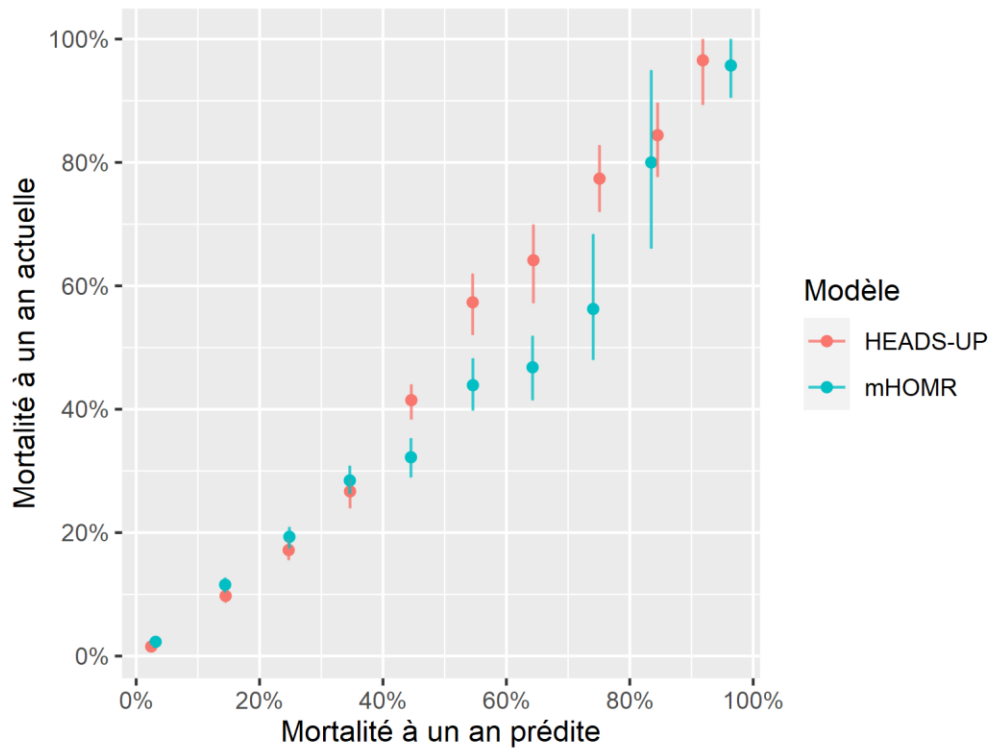


Figure 10. Pente de calibration des modèles dans l'échantillon de validation . La proportion des cas réellement décédés à l'intérieur d'un an est présentée pour chaque décile de risque prédit. Les intervalles de confiance étaient calculés avec une procédure de bootstrap à 100 itérations.

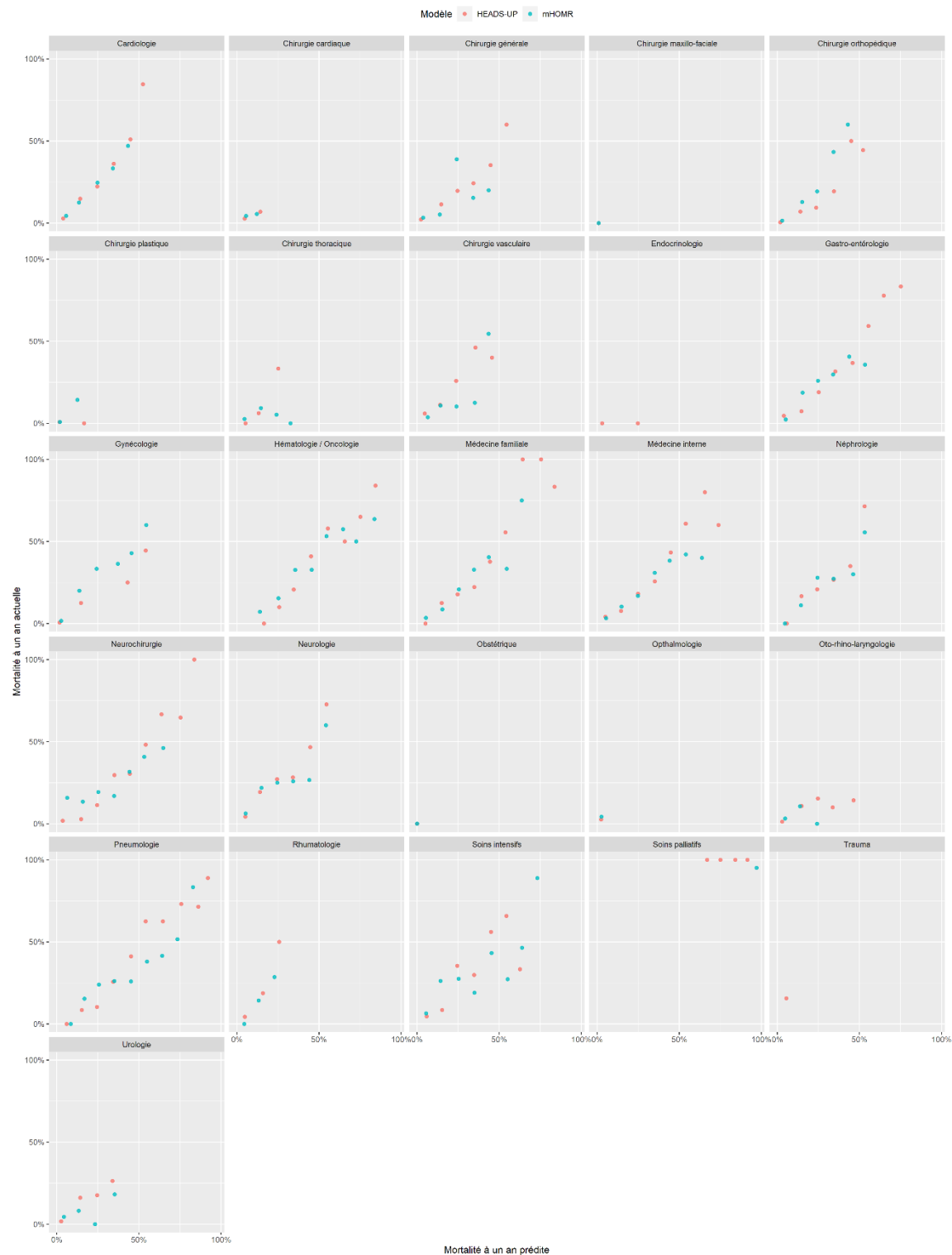


Figure 11. Pente de calibration des modèles dans l'échantillon de validation stratifié par service d'admission . Une visite par patient était aléatoirement choisie parmi les visites de chaque service. La proportion des cas réellement décédés à l'intérieur d'un an est présentée pour chaque décile de risque prédit. Les déciles avec moins de cinq observations sont exclus.

Tableau 5. Indicateurs de validité externe des modèles dans chaque service d'admission. Les indicateurs étaient calculés sur un échantillon comprenant une visite par patient aléatoirement choisie parmi les visites de chaque service. Le service d'obstétrique est exclu, car il n'y avait aucun cas de décès à l'intérieur d'un an parmi les visites choisies aléatoirement.

Service	Score de Brier		Indice de concordance	
	<i>mHOMR</i>	<i>HEADS-UP</i>	<i>mHOMR</i>	<i>HEADS-UP</i>
Cardiologie	0.0841	0.0785	0.749	0.823
Chirurgie cardiaque	0.0432	0.0402	0.551	0.633
Chirurgie générale	0.0497	0.047	0.768	0.823
Chirurgie maxillo-faciale	0.0106	0.00695	0.987	1
Chirurgie orthopédique	0.0421	0.0388	0.871	0.918
Chirurgie plastique	0.0139	0.0131	0.803	0.957
Chirurgie thoracique	0.0698	0.0591	0.55	0.687
Chirurgie vasculaire	0.0902	0.0854	0.693	0.747
Endocrinologie	0.0516	0.0456	0.844	0.959
Gastro-entérologie	0.159	0.119	0.682	0.846
Gynécologie	0.0361	0.0281	0.902	0.964
Hématologie / Oncologie	0.231	0.205	0.656	0.752
Médecine familiale	0.167	0.149	0.692	0.776
Médecine interne	0.156	0.141	0.725	0.786
Néphrologie	0.154	0.151	0.734	0.74
Neurochirurgie	0.173	0.0999	0.618	0.903
Neurologie	0.148	0.141	0.688	0.735
Ophthalmologie	0.0596	0.0441	0.695	0.93
Oto-rhino-laryngologie	0.0507	0.0482	0.79	0.86
Pneumologie	0.226	0.165	0.654	0.801
Rhumatologie	0.122	0.105	0.768	0.82
Soins intensifs	0.202	0.185	0.693	0.747
Soins palliatifs	0.0321	0.0765	0.233	0.722
Trauma	0.168	0.154	0.714	0.669
Urologie	0.0558	0.0505	0.743	0.868

La validité temporelle des deux modèles était raisonnablement similaire pour chaque mois dans la période de validation (Figure 12 et Figure 13). L'indice C médian était de 0.875 (intervalle: 0.862 - 0.890) pour la forêt aléatoire et de 0.823 (intervalle: 0.809 - 0.843) pour le modèle mHOMR. De plus, la trace de calibration semblait plus reproductible pour la forêt aléatoire que pour le modèle mHOMR (Figure 14).

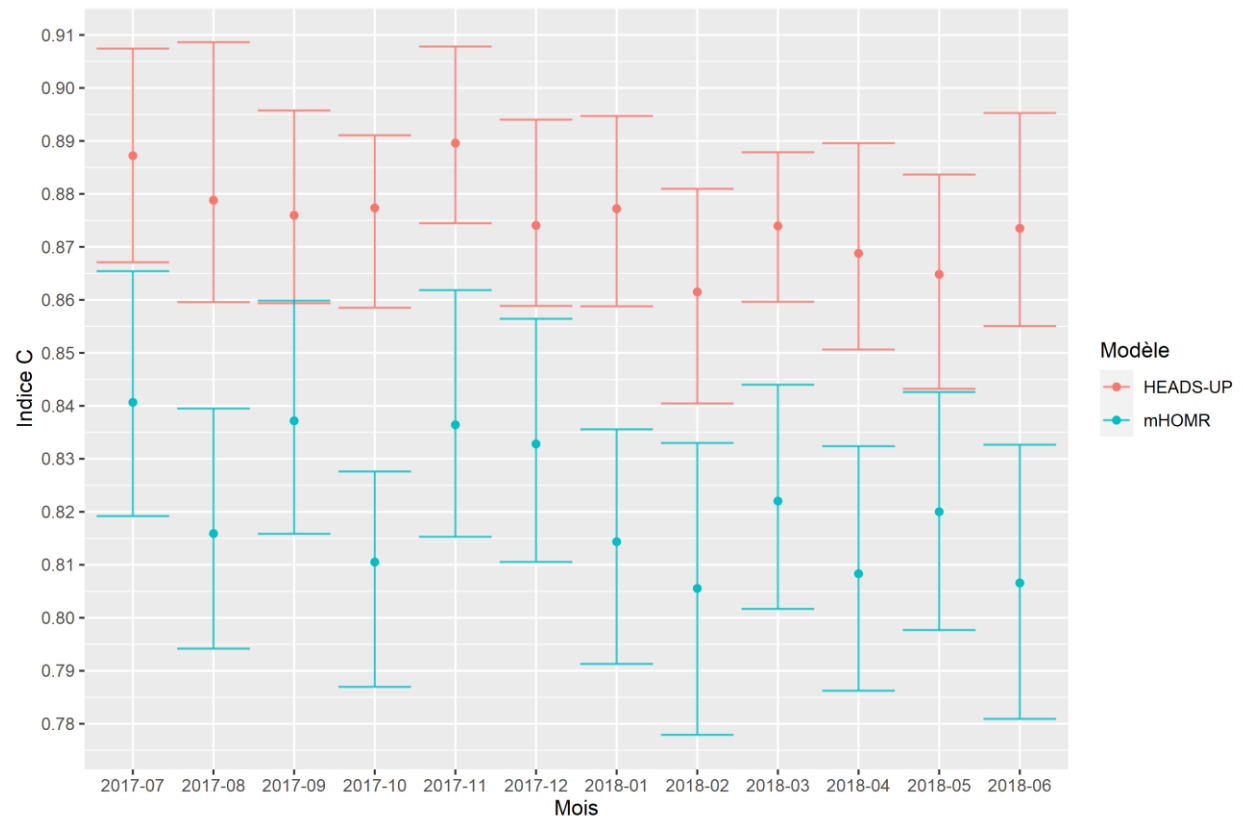


Figure 12. L'indice C en fonction du mois . L'indicateur était calculé sur un échantillon comprenant la première visite de chaque patient admis pendant le mois.

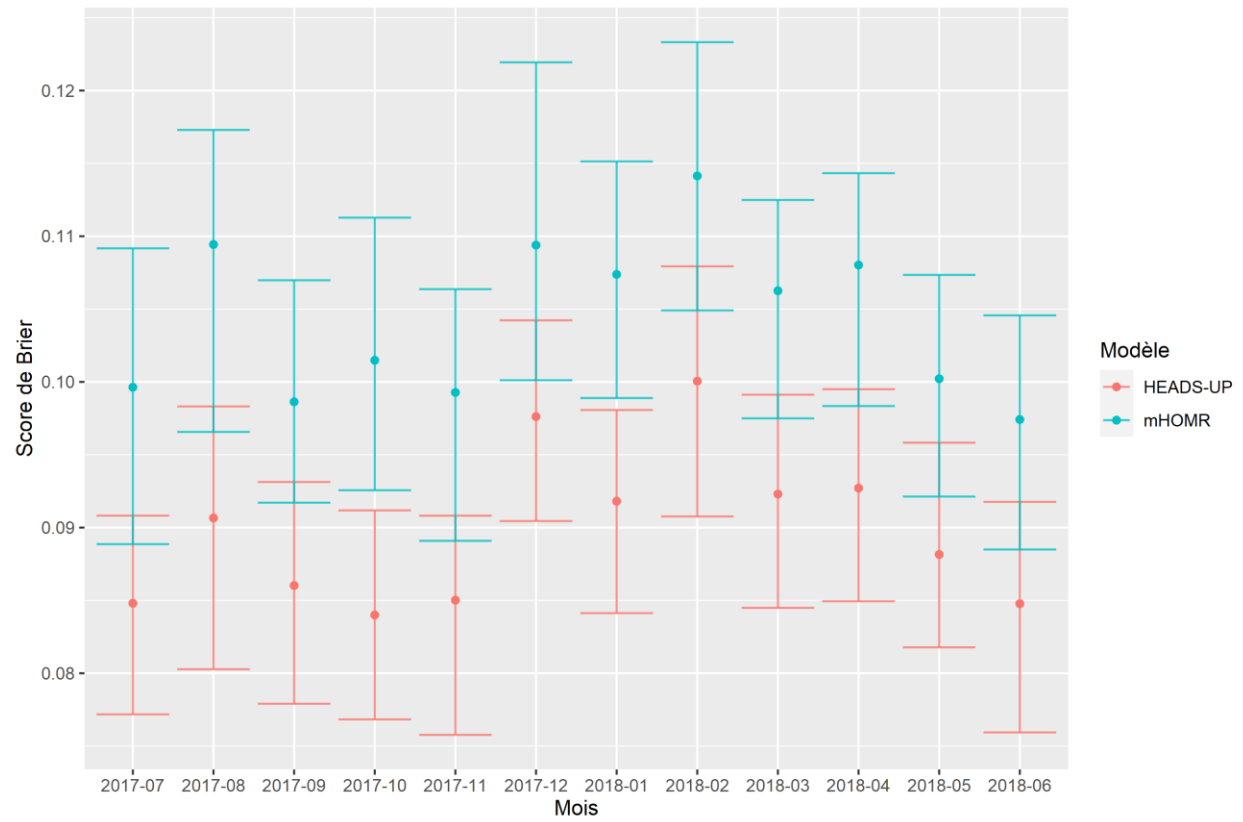


Figure 13. Le score de Brier en fonction du mois . L'indicateur était calculé sur un échantillon comprenant la première visite de chaque patient admis pendant le mois.

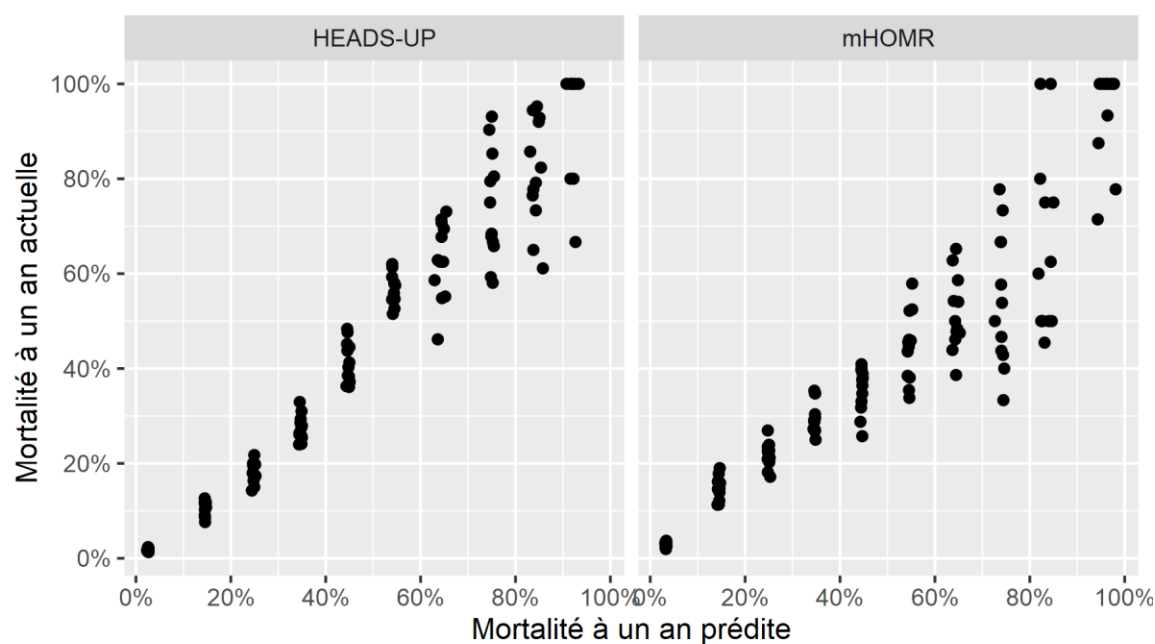


Figure 14. Pente de calibration tracée pour chaque mois . La proportion des cas réellement décédés à l'intérieur d'un an est présentée pour chaque décile de risque prédit parmi les visites de chaque mois.

4.4 Impact des stratégies de sélection de variables

La validité de HEADS-UP lorsqu'un sous-ensemble des prédicteurs finaux est inclus est présentée dans le Tableau 6. L'inclusion de soit le diagnostic d'admission ou les diagnostics antérieurs au modèle de base semblait améliorer sa validité et l'inclusion de ces deux ensembles de prédicteurs dans le modèle complet semblait améliorer la validité davantage. Une visualisation de l'importance relative des variables incluses dans le modèle HEADS-UP, utilisant la technique de décrite dans la Section 3.3, est disponible dans l'Annexe 6.

Tableau 6. Importance des variables diagnostiques dans la validité du modèle HEADS-UP. PB : Prédicteurs de base, DxAdm : Diagnostic d'admission, DxAnt : Diagnostics antérieurs

	PB	PB + DxAdm	PB + DxAnt	Tous
Indice C (IC 95 %)	0.851 (0.844 - 0.858)	0.881 (0.875 - 0.888)	0.864 (0.857 - 0.87)	0.888 (0.882 - 0.894)
Score de Brier (IC 95 %)	0.0839 (0.0816 - 0.0863)	0.0756 (0.0733 - 0.0778)	0.0807 (0.0785 - 0.083)	0.0744 (0.0722 - 0.0767)

La comparaison entre notre regroupement personnalisé des comorbidités et le regroupement du CCI pour prédire la mortalité à un an sans égard à la disponibilité des données en temps réel est présentée dans le Tableau 7 et la Figure 15. Notre regroupement était plus valide en termes de l'indice C et le score de Brier. Par contre, bien que les deux regroupements avaient une tendance à surestimer la probabilité de mortalité à un an, en regardant la pente de calibration, notre regroupement semblait surestimer cette probabilité davantage.

Tableau 7. Validité du regroupement personnalisé comparé au Charlson Comorbidity Index (CCI)

	CCI	Regroupement personnalisé
Indice C (IC 95 %)	0.875 (0.868 - 0.881)	0.920 (0.915 - 0.926)
Score de Brier (IC 95 %)	0.0785 (0.0762 - 0.0808)	0.0602 (0.0580 - 0.0623)

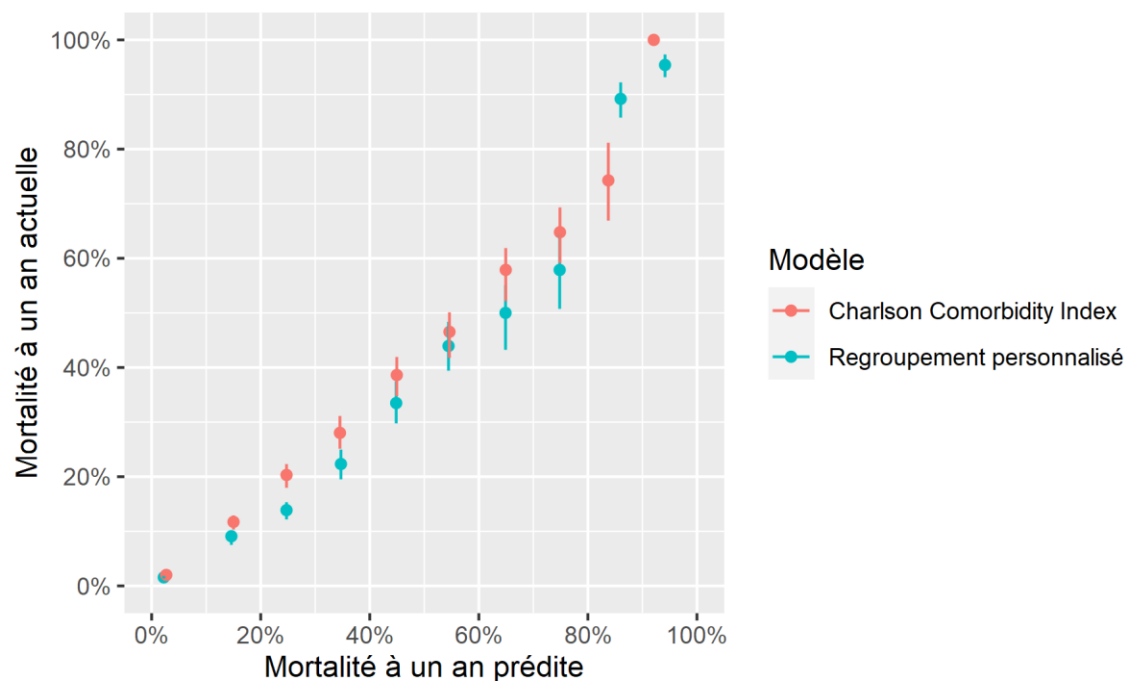


Figure 15. Calibration du regroupement personnalisé comparé au Charlson Comorbidity Index

4.5 Utilité clinique

Parmi les 26 268 hospitalisations (19 992 patients) entre le 1^{er} juillet 2017 et le 1^{er} juillet 2018, un SR était documenté pendant la visite dans 7 800 (29,7 %). Le SR était documenté soit pendant la visite ou l'année précédant la visite relativement à 9 400 hospitalisations (35,8 %; SR absent pour

64,2 % des hospitalisations). Parmi ces 9 400 hospitalisations, la dernière valeur était RCR/IET pour 3 185 hospitalisations (33,9 %), NPR/IET pour 1 137 hospitalisations (12,1 %) et NPRI pour 5 078 hospitalisations (54 %). Les caractéristiques des hospitalisations selon l'absence du SR et selon la dernière valeur documentée se retrouvent dans le Tableau 8 et le Tableau 9, respectivement.

Tableau 8. Description de l'échantillon de validation stratifié par l'issue de statut de réanimation (SR) absent, c.-à-d. aucun SR documenté pendant la visite ou l'année précédant la visite. Les variables continues sont rapportées comme médiane [écart interquartile]. Les variables catégoriques sont rapportées comme fréquence (pourcentage)

	SR absent	
	<i>Vrai</i>	<i>Faux</i>
Nombre de visites	16868	9400
Âge médian	59.00 [36.00, 70.00]	75.00 [65.00, 85.00]
Sexe = Masculin (%)	7391 (43.8)	4761 (50.6)
Type de service (%)		
Chirurgicale	8101 (48.0)	1614 (17.2)
Médicale	5577 (33.1)	7689 (81.8)
Obstétrique	3180 (18.9)	5 (0.1)
Soins palliatifs	10 (0.1)	92 (1.0)
Soins intensifs pendant la visite = Vrai (%)	1784 (10.6)	1812 (19.3)
Admission urgente = Vrai (%)	9124 (54.1)	8656 (92.1)
Mortalité à l'intérieur d'un an = Vrai (%)	920 (5.5)	3043 (32.4)
Probabilité de mortalité médiane (HEADS-UP)	0.03 [0.00, 0.12]	0.31 [0.16, 0.45]
Probabilité de mortalité médiane (mHOMR)	0.05 [0.02, 0.13]	0.27 [0.15, 0.39]

Tableau 9. Description de l'échantillon de validation selon la dernière valeur du statut de réanimation (SR) lorsque ce dernier était documenté pendant la visite ou l'année précédant la visite. Les variables continues sont rapportées comme médiane [écart interquartile]. Les variables catégoriques sont rapportées comme fréquence (pourcentage).

	Préférence documentée du statut de réanimation		
	<i>RCR/IET</i>	<i>NPR/IET</i>	<i>NPRI</i>
Nombre de visites	3185	1137	5078
Âge médiane	66.00 [56.00, 74.00]	74.00 [68.00, 82.00]	82.00 [73.00, 88.00]
Sexe = Masculin (%)	1867 (58.6)	614 (54.0)	2280 (44.9)
Type de service (%)			
Chirurgicale	771 (24.2)	212 (18.6)	631 (12.4)
Médicale	2408 (75.6)	925 (81.4)	4356 (85.8)
Obstétrique	5 (0.2)	0 (0.0)	0 (0.0)
Soins palliatifs	1 (0.0)	0 (0.0)	91 (1.8)
Soins intensifs pendant la visite = Vrai (%)	906 (28.4)	352 (31.0)	554 (10.9)
Admission urgente = Vrai (%)	2725 (85.6)	1058 (93.1)	4873 (96.0)

Mortalité à l'intérieur d'un an = Vrai (%)	399 (12.5)	309 (27.2)	2335 (46.0)
Probabilité de mortalité médiane (HEADS-UP)	0.18 [0.07, 0.32]	0.30 [0.17, 0.42]	0.38 [0.26, 0.50]
Probabilité de mortalité médiane (mHOMR)	0.16 [0.08, 0.29]	0.25 [0.15, 0.38]	0.32 [0.22, 0.43]

À un seuil de 0,210, le modèle mHOMR aurait généré 8 559 alertes entre le 1^{er} juillet 2017 et le 1^{er} juillet 2018 (23,4 alertes par jour). Le modèle HEADS-UP aurait généré le même nombre d'alertes à un seuil de 0,208. La performance de ces deux modèles, lorsqu'ils génèrent la même quantité d'alertes, est présentée dans le Tableau 10. À noter que le modèle HEADS-UP était supérieur au modèle mHOMR pour tous les indicateurs de classification testés ($p < 0.01$).

Tableau 10. Comparaison de la performance pronostique des modèles lorsqu'ils auront généré 8 559 alertes dans l'échantillon de validation. Les estimés sont présentés avec leur intervalle de confiance 95%.

	HEADS-UP	mHOMR	
Seuil	0.208	0.210	
Sensibilité	82.4 % (81.3 % - 83.8 %)	73.9 % (72.6 % - 75.2 %)	$p < 0.0001$
Spécificité	74.5 % (73.8 % - 75.2 %)	73.0 % (72.2 % - 73.7 %)	$p = 0.0002$
Valeur prédictive positive	36.3 % (35.2 % - 37.2 %)	32.5 % (31.4 % - 33.6 %)	$p < 0.0001$
Valeur prédictive négative	96.0 % (95.7 % - 96.4 %)	94.1 % (93.7 % - 94.5 %)	$p < 0.0001$

Les deux modèles permettaient d'identifier des patients avec un risque de mortalité à un an élevé (taux de mortalité à un an : 33-36 %) et des patients qui n'avaient pas eu un SR (taux de SR absent : 28-29 %). Le taux conjoint de la mortalité à un an et une absence du SR était 6,7 % et 5,1 % parmi les alertes pour le modèle HEADS-UP et le modèle mHOMR, respectivement. En général, l'absence du SR était associée avec un plus faible taux de mortalité comparativement à lorsqu'il était documenté (rapport de cotes [IC 95 %] : 0,12 [0,11-0,13]). Cette association est cohérente avec la tendance naturelle des cliniciens à préférentiellement documenter le SR chez des patients qu'ils jugent à haut risque de mortalité. Il y avait toujours une association significative entre l'absence du SR et la mortalité parmi les patients dont les modèles prédisaient un risque de mortalité qui était au-dessus leurs seuils respectifs (rapport de cotes : 0,34 et 0,47 pour le modèle mHOMR et le modèle HEADS-UP, respectivement); la persistance de cette association après

l'application d'un seuil de risque est témoin de la capacité discriminative des cliniciens avec l'information pronostique non disponible aux modèles. Cela dit, le taux de mortalité à un an était tout de même élevé parmi les patients avec un SR absent et une alerte (26 % et 20 % pour le modèle HEADS-UP et le modèle mHOMR, respectivement).

La différence entre les deux modèles est plus apparente lorsqu'ils sont analysés à différents seuils. La Figure 16 présente le nombre d'alertes par jour selon le seuil décisionnel S . Le taux de mortalité à un an et le taux d'absence du SR sont présentés selon le nombre d'alertes dans la Figure 17 et Figure 18, respectivement. Par exemple, à un seuil de 0,5, le modèle HEADS-UP a généré en moyenne six alertes par jour. Le patient est décédé à l'intérieur d'une période d'un an parmi 65 % des alertes, le SR était absent pendant la visite et l'année précédant la visite parmi 25 % des alertes, et le patient avait les deux issues parmi 15 % des alertes. Le taux de SR absent était majoritairement plus élevé pour les alertes du modèle mHOMR, mais le taux de mortalité y étant associé était toujours plus élevé pour les alertes de la forêt aléatoire. Le modèle HEADS-UP avait un bénéfice net plus élevé à tout seuil jusqu'à 0,8 selon l'analyse de courbe décisionnelle (Figure 19).

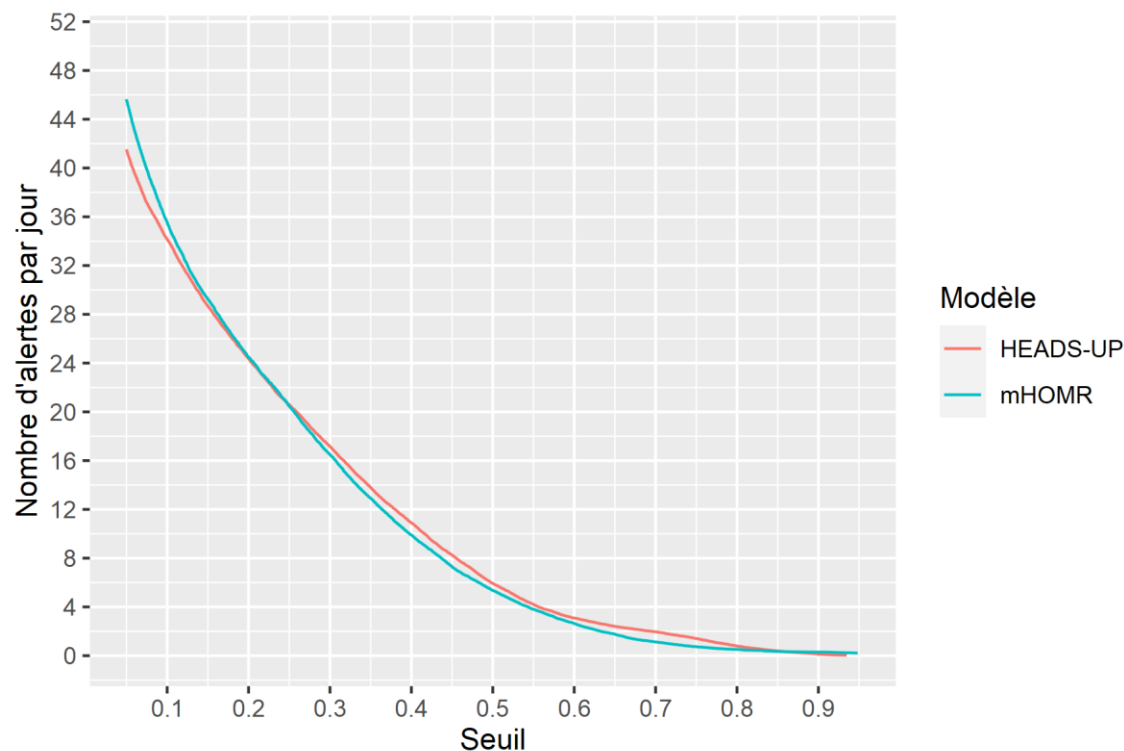


Figure 16. Nombre moyen d'alertes par jour en fonction du seuil décisionnelle

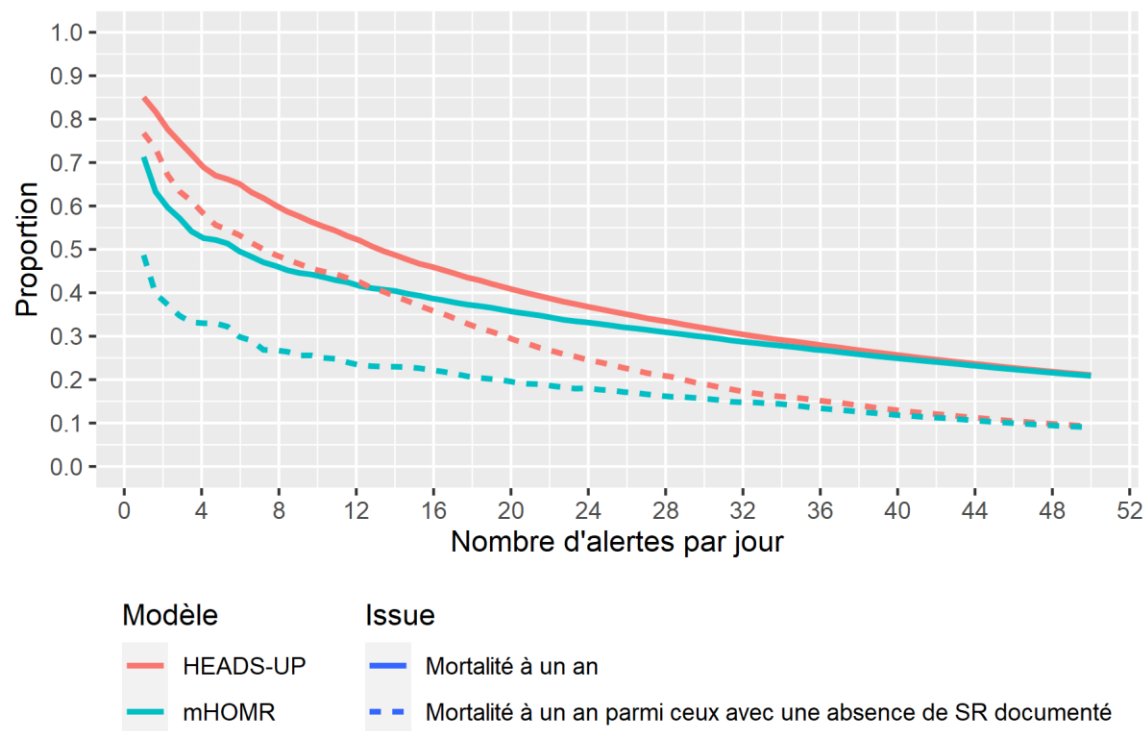


Figure 17. Valeur prédictive positive (VPP) des alertes pour la mortalité à un an en fonction du nombre moyen d'alertes par jour. La ligne solide représente la VPP parmi toutes les alertes et la ligne pointillée représente la VPP parmi les alertes où le SR est absent.

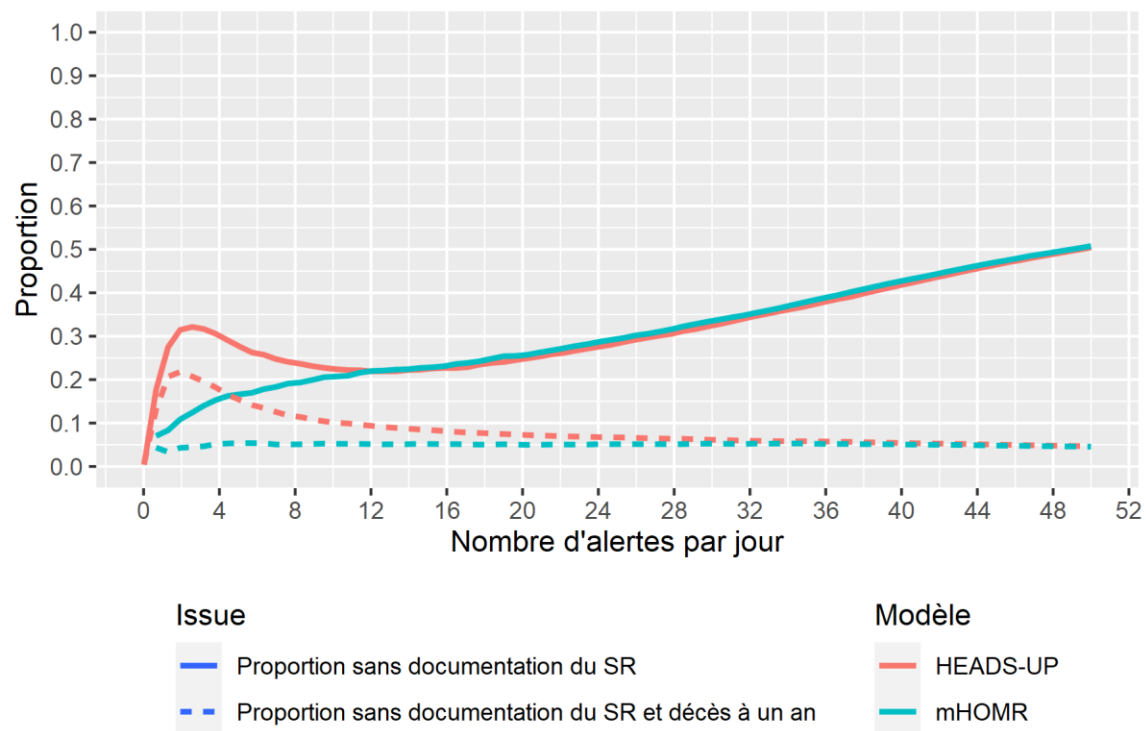


Figure 18. Valeur prédictive positive (VPP) des alertes pour l'absence du SR en fonction du nombre moyen d'alertes par jour . La ligne solide représente la VPP pour l'absence du SR et la ligne pointillée représente la VPP pour l'absence du SR et la mortalité à un an.

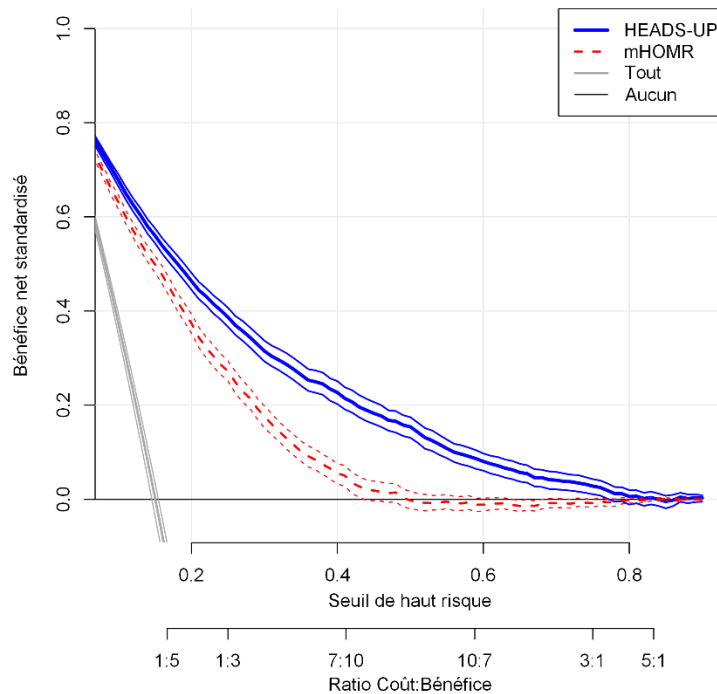


Figure 19. Comparaison du bénéfice net des modèles avec une courbe décisionnelle. Les courbes représentent le bénéfice net (axe y) d'une stratégie d'agir pour les cas au-dessus un seuil donnée (axe x) selon le modèle indiqué. « Tout » et « Aucun » correspond à une stratégie d'agir pour tous les cas et une stratégie d'agir pour aucun cas, respectivement. Cette analyse assume que le seuil de haut risque utilisé est une indication du ratio coût : bénéfice.

La Figure 20 décrit les fréquences des dernières valeurs du SR lorsque ce dernier était disponible selon le risque de mortalité prédit. Pour les deux modèles, l'intensité du SR était progressivement moindre en lien avec le risque de mortalité prédit. Parmi ceux avec un SR documenté, la majorité des patients ne voulaient pas une RCR lorsque leur risque de mortalité prédit était au-dessus de 14-15 %. De manière intéressante, ce risque transitionnel était très similaire pour les deux modèles et correspond approximativement avec la prévalence de mortalité à un an dans l'échantillon.

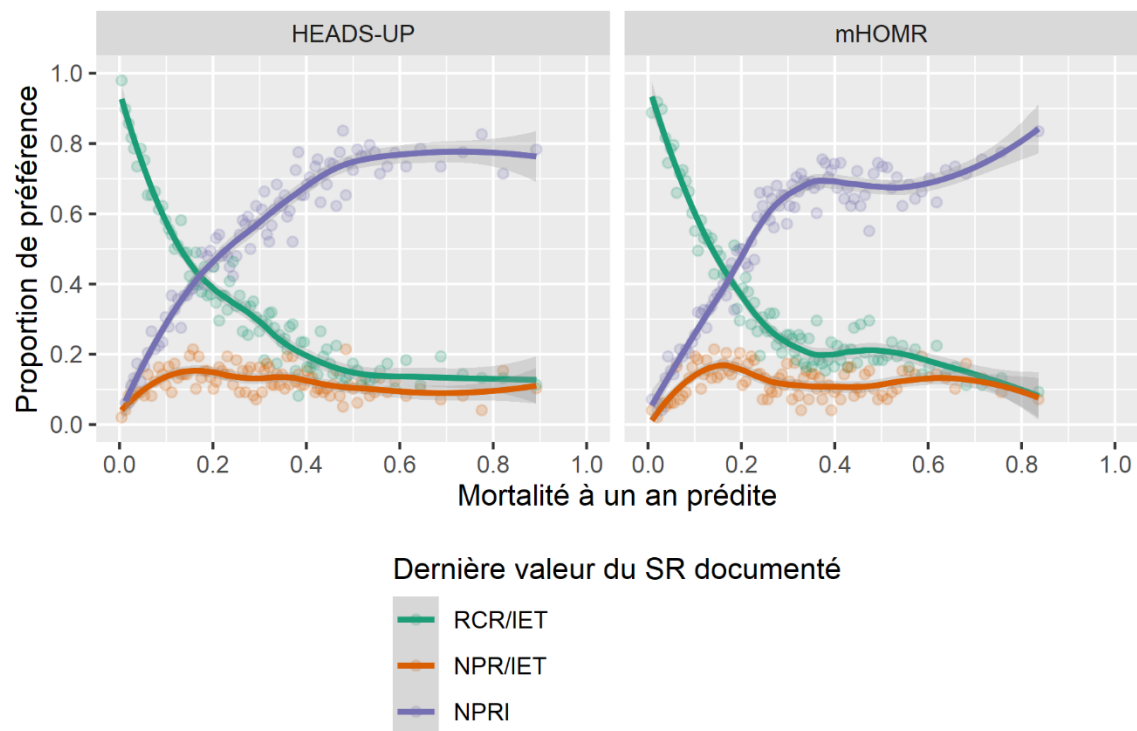


Figure 20. Dernière valeur du statut de réanimation (SR) documenté en fonction du risque de mortalité prédit de chaque modèle

Chapitre 5

Discussion

Le projet de maîtrise présenté ici visait à mettre à disposition un modèle de prédiction de mortalité chez des patients hospitalisés au CHUS et donc en étant compatible avec les systèmes informatiques dans la région de l'Estrie. Après détermination des exigences, une analyse des modèles existants n'a pas permis d'en identifier un répondant à toutes les attentes. Un modèle utilisant une approche différente (les forêts aléatoires) a donc été développé. Ce modèle répond à toutes les exigences nécessaires d'un système d'aide à la décision clinique qui identifie des patients à haut risque de mortalité afin de stimuler les discussions sur les objectifs de santé. Il a également été possible de démontrer que notre nouveau modèle est plus valide et plus utile cliniquement que le seul modèle de référence potentiellement applicable (avec quelques ajustements) au CHUS pour nos besoins.

5.1 Validation temporelle

La bonne validité de HEADS-UP dans l'échantillon de validation démontre qu'il est suffisamment robuste aux changements temporels qui ont eu lieu au CHUS entre la période d'entraînement et de validation. Une perte en validité était tout de même notée (p. ex., baisse de l'indice C de 0.91 à 0.89), probablement dû aux changements graduels dans les soins sur quelques années, des changements qui ne sont pas prévisibles au moment de l'entraînement.

5.2 Validité locale vs faisabilité multicentrique

La bonne validité externe et utilité clinique du HEADS-UP soutiennent son utilisation à notre centre dans l'avenir. Une fois qu'elle est entraînée, la puissance de calcul nécessaire pour appliquer la forêt aléatoire à chaque admission est triviale. Le travail nécessaire pour combiner les sources de données dans une SADC est légèrement plus élevé pour HEADS-UP, qui nécessite la base de données SIGDU en plus du DCI-Ariane et MED-ÉCHO. Par contre, la combinaison de ces sources était déjà effectuée pour l'entrepôt CIRESSS et elle est donc faisable. Éventuellement, la « Plateforme apprenante de santé et des services sociaux » (PARS3) [57] facilitera la combinaison de ces sources de données.

Bien que le modèle mHOMR soit au final applicable une fois les transformations requises effectuées, il ne présente tout de même pas une validité aussi élevée dans notre centre que HEADS-UP et n'identifie pas autant d'individus avec un risque de mortalité élevé et une absence de SR. Il ne présente pas non plus un profil d'interprétabilité (et donc de confiance) aussi favorable.

La discrimination de mHOMR tel qu'évaluée au CHUS n'était pas aussi bonne que celle publiée pour l'évaluation en Ontario, même lorsque les coefficients étaient redérivés et le regroupement des services révisé pour notre évaluation. Une baisse de validité a aussi été observée lorsque le modèle HOMR original a été validé en Irlande [58]. À noter, seulement les personnes de plus de 65 ans étaient incluses pour cette validation avec une cohorte relativement petite. Une baisse en validité lorsqu'un modèle de prédiction est porté hors de son centre de développement est une réalité fréquente qui reflète souvent une différence dans les caractéristiques de la cohorte de validation [19]. Par contre, notre cohorte de validation était large et les critères d'inclusion des patients étaient similaires. Une raison pouvant expliquer une telle différence se trouve au niveau de la disponibilité moindre des données informatisées sur le milieu de vie d'un patient au moment d'admission. Dans le paradigme traditionnel de recherche sur des modèles de prédiction, on développe des modèles dans un milieu et démontre qu'il est aussi valide dans d'autres milieux. Lorsqu'on développe un modèle de prédiction automatisée qui exploite des ressources informatiques, cette stratégie de recherche est sous-optimale; les exigences informatiques du dénominateur commun limiteront le domaine scientifique dans son ensemble. Si on se limite à opérationnaliser que des interventions applicables avec l'infrastructure informatique la moins développée, on ne saurait pas le coût d'opportunité de ne pas investir dans le développement de cette infrastructure. La prolifération des modèles de prédiction sans égard à la capacité des autres milieux à l'utiliser n'est pas non plus un scénario désirable, car il peut enraciner des inégalités de soins de santé sans parler de la désorganisation de la littérature. Le développeur d'un modèle de prédiction automatisée doit donc trouver le juste milieu dans le développement des modèles maximalement utile pour son centre et maximalement généralisable aux milieux de soins similaires. L'apprentissage distribué est une stratégie qui pourrait faciliter la généralisation des modèles de prédiction automatisée à plusieurs centres dans l'avenir [59], mais il convient de souligner qu'un modèle généralisable à tous n'est pas nécessairement utile pour tous [60].

En ce sens, nous avons orienté le développement de HEADS-UP pour optimiser sa validité et son utilité pour notre institution, même si certains prédicteurs, comme le diagnostic d'admission, ne sont pas universellement disponibles. En fait, l'informatisation d'un diagnostic au moment d'admission est la seule exigence de notre modèle qui n'est pas garanti pour tout centre hospitalier au Québec. L'exclusion du diagnostic d'admission rend notre modèle nécessairement plus généralisable; lorsqu'exclu, il peut être appliqué dans un hôpital qui ne l'enregistre pas dans le DCI au moment d'admission. Empiriquement, par contre, la présence de cette variable rend le modèle plus valide pour notre centre, et potentiellement pour les autres centres aussi. Si on démontre l'utilité clinique du modèle dans notre centre, il serait alors raisonnable de considérer investir pour l'implantation et la standardisation de cette variable informatique dans chaque centre (surtout que ces données seraient largement utiles pour plusieurs autres fonctions). Une restriction prioritaire des données utilisées pour favoriser principalement la portabilité des modèles de prédiction automatisée pourrait mener à un moins grand élan pour améliorer l'infrastructure et l'interopérabilité nécessaire pour appliquer des modèles avec une plus grande potentielle d'impact clinique.

5.3 Communication d'information dans un SADC

Dans notre analyse des exigences, nous avons considéré que la seule information nécessaire à communiquer aux cliniciens pour notre scénario clinique était si le risque de mortalité prédit était au-dessus un certain seuil. Nous n'avons pas inclus l'exigence de communiquer le risque estimé, avec ou sans intervalle de confiance. Élaborons sur les facteurs influençant notre analyse.

D'abord, la problématique qui justifie notre solution est que les DOS n'ont pas lieu aussi souvent qu'elles devraient. Parmi les causes de cette problématique est l'incertitude du pronostic tel que perçu par les cliniciens, mais clarifions qu'il est tout à fait acceptable pour un clinicien d'amorcer une DOS malgré cette incertitude ; le rôle du clinicien n'est pas de prendre une décision à propos des objectifs de soins voulue par le patient, mais d'apporter au patient les éléments qui informent cette décision, incluant toute incertitude. Il n'y a pas d'évidence qu'une DOS, jugée pertinente par un clinicien, pourrait être néfaste s'il existe une incertitude dans le pronostic [61]. À cet effet, à notre avis, le rôle principal d'un SADC dans notre scénario clinique n'est pas autant de préciser le pronostic exact, mais de surmonter la barrière comportementale que le sentiment d'incertitude génère parmi les cliniciens [5]. Sur cette base, il n'est pas nécessaire de présenter le risque exact

de mortalité à un an, seulement un indice qu'un DOS pourrait être bénéfique, ce qui est atteint par un indicateur dichotomique. Si l'espace de décisions que le SADC est censé appuyer se grandit, alors l'information communiquée doit être ré-évaluée. Par exemple, si on ajoute l'objectif d'augmenter les consultations appropriées en soins palliatifs, il peut devenir souhaitable de communiquer un niveau plus granulaire de risque de mortalité, car le seuil décisionnel pour amorcer une DOS n'est pas le même que celui pour une consultation appropriée en soins palliatifs.

Il pourrait être souhaitable d'offrir plus d'information à un clinicien dans certains contextes^{§§}, mais il faut que le clinicien sache quoi faire avec cette information supplémentaire [63]. Nous n'avons pas d'évidence présentement sur la façon dont le risque exact (selon un modèle donné) devrait être utilisé dans le contexte d'une DOS routinière. La surcharge cognitive peut réduire la sécurité des soins [64] et il y a une reconnaissance croissante dont l'information présentée à des cliniciens dans les SADC doit être justifiée par les ressources d'attention qu'elle consomme [65]–[67]. Comme développeur d'un SADC, si on ne sait pas comment un certain élément d'information devrait être utilisé, il est probable que les cliniciens ne sachent pas non plus, et il est possible que cet élément augmente la surcharge cognitive sans apporter un bénéfice sur les issues cliniques visé par le SADC. Par conséquent, notre perspective est que l'information communiquée à des cliniciens dans un SADC doive être le minimum nécessaire pour atteindre les objectifs du système : dans notre cas, maximiser le nombre de DOS parmi les patients qui sont à la fin de vie. Un élément d'information donné, comme une estimation de point ou un intervalle de confiance, pourrait être inclus si jamais un bénéfice potentiel en lien se spécifie, mais tout excès d'information sans justification est une source de risque dans un contexte où la capacité des humains à gérer l'information est déjà maximisée – d'où la nécessité du SADC.

5.4 Forces

Pour le design du modèle HEADS-UP, la méthode de la forêt aléatoire a été utilisée comme méthode d'apprentissage sur des données administratives informatisées, incluant du texte non structuré, qui sont disponibles au moment de l'admission. Cette méthode d'apprentissage, cette source de données et ce type de données pour prédire la mortalité à un an chez des patients hospitalisés sont des éléments qui ont tous été décrits comme pertinents dans la littérature

^{§§} Par exemple, si le but est de trier des ressources limitées [62], ce qui n'est aucunement l'intention de ce travail.

récente [30], [34], [34], [39], [40]. En plus de combiner ces caractéristiques, le travail de recherche qui est présenté dans le cadre du présent mémoire contribue à la littérature scientifique portant sur les modèles de prédiction. La contribution s'inscrit selon les deux éléments principaux ci-dessous ont été démontrés :

1. L'inclusion des données hautement dimensionnelles avec des stratégies qui maintiennent davantage d'interprétabilité comparativement à l'apprentissage profond ou la régression logistique avec intrants fortement modifiés.
2. L'évaluation de l'utilité clinique du modèle au-delà des indicateurs de classification usuelle.

5.4.1 Interprétabilité avec des données hautement dimensionnelles

Pour inclure un diagnostic d'admission non structuré comme prédicteur, nous avons manuellement conçu une correspondance vers des termes reconnaissable pour les cliniciens. À notre connaissance, nous sommes les premiers à démontrer la faisabilité de cette approche qui augmente la validité du modèle sans perte à l'interprétabilité (spécifiquement, à la confiance des utilisateurs à identifier une erreur). Des travaux récents dans le domaine de prédiction clinique utilisent des méthodes automatisées et l'apprentissage profond pour inclure le texte libre [40], [68], [69]. L'avantage principal de l'automatisation est qu'un expert de domaine n'est pas aussi nécessaire au départ, diminuant l'effort nécessaire au moment de développement. Le désavantage principal est la perte en interprétabilité, nécessitant l'effort post-développement pour augmenter la confiance des utilisateurs. À la base, si un prédicteur fourni au modèle n'est pas compréhensible à un utilisateur, soit en forme d'entrée ou en fonction de pronostic, l'utilisateur ne sera pas capable de vérifier la présence d'une erreur lorsque cette dernière est due à ce prédicteur. En plus d'améliorer l'interprétabilité, la correspondance manuelle par un expert de domaine permet de réduire des dimensions d'information superflue; si la connaissance du domaine est bonne, moins d'observations sont nécessaires pour l'entraînement du modèle. En plus de faciliter l'accès aux données de différentes sources, PARS3, qui est basé sur les ontologies, favorisera l'annotation sémantique des données et pourrait ainsi favoriser l'entraînement efficace des modèles de prédiction hautement dimensionnelle et interprétable.

Une approche alternative à la correspondance manuelle avec des expressions régulières est d'appliquer le traitement automatisé de langue naturelle (TALN) au corpus du texte libre puis

fournir au modèle un ensemble de termes contrôlés [70]. Nous n'avons pas utilisé le TALN, car il était jugé non faisable pour les termes et acronymes français (voir québécois !) hautement spécifiques au domaine clinique (québécois). Les ressources disponibles pour le TALN en médecine sont surtout conçues pour l'anglais [71]. Idéalement, un texte aussi court et cliniquement important que le diagnostic d'admission devrait être codifié par une terminologie préexistante au moment de saisie, rendant le texte structuré. Néanmoins, notre approche de correspondance entre du texte non structuré et des termes cliniques reconnaissables avant de les fournir à un modèle de prédiction nous semble généralisable dans le cas où la codification des termes à leur source n'est pas disponible. Notre processus de création des termes n'est pas déterministe, mais la terminologie finale avec les règles de correspondance est entièrement vérifiable par d'autres experts du domaine clinique et modifiable au besoin. L'utilisation éventuelle du TALN dans ce contexte est complémentaire; les termes candidats sont répertoriés, et les correspondances sont créées manuellement ou d'une manière semi-automatisée afin de regrouper des synonymes cliniques et distinguer les homonymes qui ont une sémantique cliniquement distincte selon le contexte. Une limite à prendre en compte pour notre approche est le nombre de termes étant donné l'exigence d'une partie du traitement en mode manuel.

Une pratique courante [18] dans le développement des modèles de prédiction clinique est l'inclusion des scores de comorbidité unidimensionnelle, comme le CCI. Le CCI est particulièrement problématique pour un modèle de prédiction personnalisé; rappelons que l'asthme (J45) est regroupé avec le MPOC (J44) et le cancer de thyroïde (C73) est regroupé avec le cancer de pancréas (C25) [72]. Les auteurs du CCI ont conçu le score pour l'ajustement du risque dans des études épidémiologiques et non la prise de décision individuelle. Nous soutenons que le regroupement des diagnostics trop dissimilaire par rapport au pronostic n'est pas une bonne pratique; en plus d'être difficile à vérifier lorsque « aplatie » en une variable continue, il diminue la confiance des utilisateurs.

Au lieu d'inclure ce genre de regroupement, nous avons fait notre propre regroupement des comorbidités dirigé par un expert de domaine. Comme la correspondance du diagnostic d'admission, nos règles de regroupement ne sont pas déterministes, mais ils sont transparents et entièrement vérifiables. Le fait d'observer que notre regroupement est plus valide, au moins en

termes de discrimination, que le regroupement du CCI dans un contexte épidémiologique, pour lequel la CCI est conçue, est un appui empirique à notre choix de regroupement.

5.4.2 Évaluation d'utilité clinique

Nous avons évalué les utilités cliniques de HEADS-UP et de mHOMR en étudiant le taux historique de documentation du SR chez les patients à haut risque de mortalité selon chaque modèle. L'étude rétrospective des processus cliniques liés à l'utilité est une stratégie connue pour estimer l'impact d'un modèle de prédiction avant de l'appliquer [42]. Il y a relativement peu d'études qui évaluent l'utilité clinique en plus de la validité d'un modèle de prédiction [13]. À notre connaissance, nous sommes les premiers à décrire l'utilité d'un modèle de prédiction de mortalité en évaluant le taux de documentation historique des objectifs de soins (au moins une partie dans notre cas). Cette évaluation n'est pas menée pour remplacer une étude d'impact prospective, mais elle peut justifier la dépense de ressource nécessaire pour implanter une SADC puis pour procéder avec une telle étude.

Nos résultats suggèrent que HEADS-UP est cliniquement utile. À un seuil de 0,208, HEADS-UP identifiait en moyenne 24 patients par jour dont 36 % décédaient à l'intérieur d'un an et 28% n'ont pas eu une documentation du SR, une partie clé du DOS, pendant l'hospitalisation en question ni l'année précédant l'hospitalisation.

5.5 Limitations

5.5.1 Absence des données paraclinique

Notre modèle n'inclut pas des données portant sur les résultats de laboratoire, pathologie, radiologie et autres investigations pertinentes pour la pronostication (p. ex. tests de fonction respiratoire) qui sont disponibles dans divers systèmes informatiques. Ces variables étant couramment utilisées par les cliniciens pour stratifier le risque de mortalité, on estime que notre modèle aurait été plus valide s'ils étaient inclus. Cependant, leur extraction pour un échantillon aussi large n'était pas faisable avec l'infrastructure informatique actuelle du CHUS. C'est aussi évidemment un enjeu pour les autres sites cliniques québécois. Avec le choix du Québec d'implanter un système d'information des laboratoires cohérent à travers la province, les prochaines années permettront possiblement de finalement pouvoir intégrer ces données dans un modèle comme HEADS-UP.

S'il devenait accessible en temps réel, notre stratégie d'inclusion aurait reflété notre stratégie pour les diagnostics en texte libre et les diagnostics codifiés, c. à. d. la génération par un expert de domaine des variables reconnaissable au clinicien et plausiblement lié au pronostic. Les anomalies de laboratoire étant souvent transitoires (p. ex. une montée de créatinine), il sera nécessaire d'inclure une composante temporelle dans le modèle, qui n'était pas explicitement considéré dans HEADS-UP. Une stratégie potentielle avec la forêt aléatoire serait l'agrégation des résultats dans des périodes temporelles discrètes (p. ex. une variable pour la créatinine moyenne dans les derniers 30 jours, une pour les 30-90 jours, etc.), avec soit une variable indiquant l'absence des données ou une stratégie d'imputation de données manquantes (qui doit maintenir l'interprétabilité). Une approche similaire pourrait être envisagée pour d'autres types de données cliniques comme la médication.

5.5.2 Qualité des données disponibles lors de l'utilisation du modèle

Les données disponibles lors de l'utilisation projetée du modèle (en début d'hospitalisation) pourraient certainement être de meilleure qualité.

Le diagnostic d'admission en texte libre trouvé dans le DCI est créé à partir de la transcription par un commis administratif du diagnostic non structuré écrit par le médecin sur le papier d'admission. Les erreurs de transcription de diverses causes diminuent la performance de cette variable et augmentent l'effort nécessaire pour le rendre utilisable. Si le diagnostic d'admission était saisi directement par le médecin qui admet le patient, et si ce diagnostic était au moins partiellement structuré (p. ex. avec le vocabulaire SNOMED-CT), notre modèle aurait été plus facilement reproductible et portable à d'autres langues et milieux.

Le milieu de vie avant l'admission et les comorbidités documentées dans les derniers 6 mois sur la feuille sommaire n'étaient pas directement disponibles. Plusieurs stratégies étaient utilisées pour les inférer avec des données historiques, mais la performance et la reproductibilité de notre modèle sont moindres que si elles avaient été disponibles directement. Idéalement, le milieu de vie serait saisi au moment d'admission pour les admissions urgentes (il est presque toujours pertinent pour le médecin traitant) et la feuille sommaire serait remplie directement dans un format électronique. Même si les diagnostics n'étaient pas d'emblée codifiés, ils pourraient être utiles par un modèle de prédiction, tel qu'exemplifié par l'utilisation du diagnostic d'admission texte libre par notre modèle.

5.5.3 Population captive

La population de l'Estrie étant relativement captive au CHU de Sherbrooke, on dispose des données longitudinales qui peuvent être plus fragmentées dans d'autres régions. Ceci est un avantage pour l'application de notre modèle localement, mais peut limiter la généralisabilité de nos méthodes dans une région où les données historiques ne sont pas autant disponibles par un seul centre. Lorsque les données seront plus facilement partagées entre milieux de soins, par exemple avec PARS3, les silos de données seront moins une barrière au partage des modèles de prédiction clinique.

5.5.4 Indice de concordance

Nous avons rapporté l'indice C afin de comparer les modèles entre eux. Néanmoins, il est délicat d'utiliser cet indicateur sans préalablement analyser les modèles comparés. Il n'est pas particulièrement étonnant qu'un modèle soit capable de discriminer la mortalité entre une patiente admise en obstétrique et une patiente admise en soins palliatifs, deux extrêmes en termes de mortalité moyenne. Dans notre contexte, l'indice C est donc certainement augmenté en lien avec l'inclusion du service d'admission.

D'un côté, la comparaison de notre modèle avec HOMR et avec mHOMR en utilisant l'indice C est probablement pertinente, car les trois modèles utilisent le service d'admission en entrée. Par contre, la comparaison avec d'autres modèles qui ne l'utiliseraient pas pourrait être trompeuse. Nous avons donc fait état de la validité et l'utilité du modèle stratifiées par service afin d'expliciter cet effet.

5.5.5 Portabilité

Notre modèle a été validé pour être appliqué au CHUS avec les données du CHUS. La validation externe temporelle nous permet d'appliquer le modèle avec confiance au CHUS, mais n'est pas suffisante pour l'appliquer dans un autre centre hospitalier. Comme décrit ci-haut, nous avons tenté de balancer le besoin d'application à plus grande échelle et le besoin d'un modèle utile. Afin d'être appliqué dans un autre centre hospitalier, il faut adapter les variables à celles disponibles dans les systèmes informatiques locaux puis revalider le modèle avec les données disponibles. L'adaptation du modèle HOMR à plusieurs reprises (HOMR, HOMR-Now!, mHOMR) dans différents centres hospitaliers démontre que cette limitation, qui est due à la disponibilité hétérogène des données de

santé, est la règle plutôt que l'exception avec ce type de modèle. Aussi longtemps que la disponibilité des données de santé n'est pas uniformisée, le développement des modèles de prédiction informatisés qui sont simultanément valides et portables ne semble pas être un objectif qui soit aisément atteignable.

5.5.6 Validation interne minimale

Pour déterminer les valeurs des hyperparamètres, nous avons utilisé un sous-échantillonnage simple de l'échantillon d'entraînement. Pour évaluer la validité interne du modèle de prédiction, nous avons utilisé une validation croisée à dix plis. Il est important de souligner qu'il existe des méthodes plus robustes pour l'estimation de la validité interne, par exemple, la validation bootstrap [11, p. 107], et que nos méthodes ne sont pas appropriées si les valeurs des hyperparamètres sont elles-mêmes issues d'une validation croisée. Pour à la fois optimiser les valeurs des hyperparamètres et évaluer la validité interne d'un modèle, il est souhaitable d'utiliser une méthode qui rend ces analyses indépendantes, comme la validation croisée « nested » [73].

Nous estimons nos méthodes juste pour notre étude en raison de la taille d'échantillon grande [74] et le paramétrage minimal de la forêt aléatoire. Notre choix du paramètre *ntree* (section 3.8.2) n'était pas basé sur une valeur optimale, mais une valeur qui permettait une validité raisonnablement proche à sa limite asymptotique théorique [16, p. 596] avec les ressources de calcul à notre disposition (et avec la valeur de *mtry* à sa valeur par défaut). Le paramètre *mtry* (section 3.8.3) était basé principalement sur la valeur suggérée par défaut, une hypothèse à priori que notre expérience a appuyée, et non un choix basé sur la valeur optimale dans les expériences de validation. Nous n'avons pas tenté d'optimiser d'autres hyperparamètres, par exemple, la profondeur des arbres, spécifiquement en raison du risque de surajustement et le désir d'avoir un modèle plus facile à reproduire.

Une baisse en validité similaire entre la validation interne et externe du modèle HEADS-UP et le modèle mHOMR (différence entre les indicateurs de la validité interne dans le Tableau 3 et les indicateurs pour le même modèle dans Tableau 4) est un argument que notre choix de méthodes n'a pas mené au surajustement de la forêt aléatoire avec l'échantillon d'entraînement. Par contre, réitérons que ces méthodes ne sont pas recommandées pour plusieurs contextes de développement de modèle de prédiction, soit lorsque le nombre d'observations est petit relativement au nombre de prédicteurs, ou la méthode d'apprentissage est plus à risque de surajustement. Le risque

d'utiliser nos méthodes dans de telles situations est d'avoir des estimations de validité interne trop optimiste, et de découvrir une baisse de performance exagérée lorsque le modèle est appliqué hors de l'échantillon d'entraînement.

Conclusion

L'objectif du présent travail de recherche était de développer un modèle de prédiction de mortalité chez des patients hospitalisés qui soit compatible avec les systèmes informatiques existants au CHUS et qui satisfait les exigences d'un SADC qui identifie les patients à haut risque de mortalité afin de stimuler et soutenir une DOS. Dans ce mémoire, nous avons décrit l'état des connaissances en lien avec les modèles de prédiction clinique, nous avons établi une liste des exigences basées sur ces connaissances, nous avons décrit le processus de développement d'une solution (HEADS-UP) qui satisfait ces exigences et nous avons fait état des propriétés de cette solution qui permet de vérifier la satisfaction des exigences et qui permet d'évaluer son importance relative à la solution de référence la plus applicable dans la littérature (modèle mHOMR).

La prochaine étape pour le modèle HEADS-UP est de l'opérationnaliser dans un SADC avec les diverses sources de données disponibles en temps réel. Un format de présentation approprié doit être conçu, un mécanisme de communication doit être choisi et un seuil décisionnel doit être déterminé en consultation avec les divers services médicaux en fonction du nombre d'alertes gérables. Après l'implantation du SADC, une étude prospective qui valide son impact clinique est prévue.

Annexe 1 Diagnostic d'admission

Table de correspondance entre chaque variable booléenne (colonne « Variable ») dans l'ensemble « diagnostic d'admission » inclut dans le modèle HEADS-UP et l'expression régulière qui était censée identifier ce diagnostic dans le texte libre saisi au moment d'admission (colonne « RegEx texte »). Pour certains diagnostics, une deuxième correspondance était faite avec une expression régulière représentant un code CIM-9 qui pouvait être dans le texte libre (colonne « RegEx CIM-9 »). Fichier disponible sous forme numérique au format CSV sur demande (annexe_1_dx_admission.csv).

Variable	RegEx texte	RegEx CIM-9
Épanchement péricardique	epanchement(?:.+?)pericard	NA
Épanchement pleurale	\bepanchement(?:.+?)(pleural pleureux)	511[.][0-9]
Étude électrophysiologique	(e p s)(electrophysio)(\beps\b)\bdefib\bpace(?:.+?)defib\bpace(.)?maker\bpace\b(\bpmp\b)	NA
Abcès	(abces) abcde collection	NA
Alcool	\broh\b alcohol ethyl delirium tremen	NA
Amygdalite	amygdalite amygdallect	NA
Anémie	anemie anémique	285[.][9] 280[.][9]
Anevrisme	anevrism anevrism	NA
Anevrisme de l'aorte	\baaa\b anevrism(?:.+?)aorte	441[.][4]
Angine	(angine)(angor)\bmcas\b maladie(?:.+?)cardiaque	411[.][90] 413[.][429[.][21] 414[.][0]
Anomalie de conduction	bloc(?:.+?)([23]e complet\b ba[. -]?v\b auricul)\bb[.][.]?a[.][.]?v\b mobitz bradycard\b bb[.][.]?g\b (\bsick sinus\b)(\bss\b) maladie(?:.+?)sinus(\bpause(s)?\b)(tachy(?:.*?)brady)	426[.][0-689]
Appendicite	app[ea]ndicit appendicect appendiculaire	54[013][.][9]
ARDS	\bards\b	NA
Arrythmie	arythmi arrythmi(\bve\b)(cardioversion)	427[.][479] 426[.][7]
Arthropathie	arthro tendon articula coxalgie ankylose	715[.]
Ascite	\basцит	789[.][5]
Asthme	asthm[ea] broncho(\W)?spas[tm]	493[.][519.1]
AVC	(\bavc\b)(a v c)(acc(?:.+?)cerebr(?:.+?))(\bavc\b)(a(\W)?v(\W)?c)\bavc\b acc(?:.+?)vasc(?:.+?)cere(?:.+?)	436[.][9] 784[.][3] 342[.]
Bronchiectasie	bronchiect	NA
Bronchite	bronchite bronchiolite surinfection bronchiq	NA
C. Difficile	\bc(?:.+?)diff	NA
Céllulite	\bcellulit	682[.][67]
Cancer	(cancer)(\badenoca)(\badenok)(\bneo)(\bad(n)?k\b) neoplasie bcarcino anaplasie sarcome(\bca\b) hodgkin(\blnh\b)(lymphome)\blma\b melanome leucemie	NA
Cancer coloréctale	((cancer)(\badenoca)(\badenok)(\bneo)(ad(n)?k)(\bca\b))(?:.+?)(coliq \bcaec sigmoid\b colo rect appendice\b grele\b)	15[34][.]
Cancer d'oesophage	((cancer)(\badenoca)(\badenok)(\bneo)(ad(n)?k)(\bca\b))(?:.+?)(oesophag)	150[.][9]
Cancer de cerveau	(\bgbm\b)(glioblasto) astrocytom\b gliome	191[.]
Cancer de foie	cholangiocarc carcinome hepat hepato(-?)carcinome hepatome\b chc\b	NA
Cancer de l'estomac	((cancer)(\badenoca)(\badenok)(\bneo)(ad(n)?k)(\bca\b))(?:.+?)(gastr estomac)	151[.][9]
Cancer de pancréas	((cancer)(\badenoca)(\badenok)(\bneo)(ad(n)?k)(\bca\b))(?:.+?)(pancrea)	157[.][9]
Cancer de poumon	((cancer)(\badenoca)(\badenok)(\bneo)(ad(n)?k)(\bca\b))(?:.+?)(poumon pulmon\b si[gd]\b lobe (inf sup moy))	16[23][.]
Cancer de prostate	((cancer)(carcinome)(\badenoca)(\badenok)(\bneo)(ad(n)?k)(\bca\b))(?:.+?)(prostat) prostact	185[.]
Cancer de sein	((cancer)(carcinome)(\badenoca)(\badenok)(\bneo)(ad(n)?k)(\bca\b))(?:.+?)(bsein\b mammaire canalaire lobulaire)	174[.][9]
Cancer de thyroïde	((cancer) carcinome(\bneo)(masse) nodule tumeur(\bca\b))(?:.+?)(thyroid folliculaire papillaire) thyroïdect	193[.]
Cancer de vessie	((cancer) carcinome(\bneo)(masse)(\bca\b))(?:.+?)(vessie vesical urothelial ureteral)	NA
Cancer ORL	((cancer)(\badenoca)(\badenok)(\bneo)(ad(n)?k)(\bca\b))(?:.+?)(\borl\b larynx pharynx langue)	161[.]
Chimiothérapie	(\bcotx\b)(chimio)(\bca\b)	NA
Chirurgie bariatrique	obesite(?:.+?)(severe morbide)	NA
Choc	\bchoc\b(?: ana)(?! hem)(?! hypo)	785[.][5]
Cholécystite	cholecystit cholecystect	575[.]
Cholédocholithiase	choledochobcp[re]\b calcul lithias(?:.+?)(choledo)	574[.][5]

Cholangite	cholangit	576[.]
Chute	\bchute(s)?\b t(?:.+)\marche\b	781[.]2
Cirrhose	\bcirrho varice(?:.+)\oeso hemorh(?:.+)(varic oeso)\bencephalop(?:.+)\hepat	571[.]5 572[.]21 456[.]1
Colique biliaire	colique(?:.+)\biliair cholelithias cholestase	574[.]2
Colite	\bcolite pancolite	NA
Convulsions	status(?:.+)(epile convuls partiel)(epilep(s i))	780[.]3 345[.]9
Cytopénie	pancytopeni bicytopeni	NA
Délirium	(\bdelire\b) (\delirium) confusion(((alteration perte etat)(?:.+)\conscience) \bagitation \baec\b	293[.]3 307[.]9 348[.]3 331[.]0
Démence	\bdemenc alzheimer(\bt(?:.+)(cognitif comportement))\ btnc\b \berrance\b	29 04[.]3 310[.]1
Désordre électrolytique	(hypo hyper)(na ca k mg mag)\deshydrat	276[.]1 276[.]0 276[.]8 276[.]7 275[.]42
Désordre de l'œil	oeil \bretin ophthalm	NA
Désordre de l'oreille	mastoidit vertige \boto labyrinth cholesteatom neuron(?:.+)\vestibul \bsurdite\b \bmeni ere\b	38 0-9[.]7 780[.]42
Désordre gynécologique	\badenomyos ligature \bendometrios endometrite \bhysterect	62 56[.]3 620[.]2 617[.]0 621[.]3
Diabète	diabet(\bd[bm]\b) hyperglycemie hypoglycemie etat hyper[]?(gly osm)	250[.]09 251[.]2 790[.]6 250[.]
Dialyse	irct irc terminal \bdialyse	NA
Diarrhée	(diarrh)	NA
Discopathie	hernie(?:.+)(disc [dls]\d) discectomie discoïdect discopathie discarthro radiculopathie lombo-sciatalgie	722[.]7 724[.]3
Douleur abdominale	(douleur dlr)(?:.+)(abdo epigast)	789[.]0
Douleur au poitrine	\bdrs\b douleur(?:.+)(thorax retro(W)?stern)	786[.]5 01
Dysphagie	\bachalasie \bdysphagi	787[.]2
Dyspnée	\bdyspne	799[.]0 786[.]03
Endocardite	endocardit	424[.]9 421[.]01
Entérite	enterit \bileite	NA
Fertilité	fertilite	NA
Fibrillation auriculaire	(\bfa(p)?\b) (\bflutter\b) (\bfibri(?:.+)\aur\w+)	427[.]3
Fibrose pulmonaire	(fibrose)(?:.+)(pulm)\ bfp\b	515[.]
Fracture	(fractur) (\bfx\b) (\b[#]\b) rofi	NA
Fracture de hanche	([#] fx fracture rofi)(?:.+)(hanche femur femoral troch intertro (bas)(?:.+)\cervic peri(-)?pro (sous)(?:.+)\capi)	NA
Fracture de membre inférieur	([#] fx fracture rofi)(?:.+)(cheville tibia rotule mal(l)?(col[.]?)) perone jambe genou pied patella maisonneuve)	NA
Fracture lombaire ou pelvien	([#] fx fracture burst)(?:.+)(pelvi pubi \bacetab bassin \bl(W)?[1-5] lombaire sacrum)	80 58[.]
Gangrène associé au MVAS	NA	785[.]4
Gastrite	gastrit	535[.]5
Grossesse	travail((cesarien) (grossess) (\d+ semaine) (rupt\w+ prematu\w+) (\brppm\b) (\rciu) (placenta) (accouche) (post([-])?partum) (gestation) fertilite	6 3-7[.]0-9[.]vV 22[.]2
Hémoptysie	hemoptys	786[.]3
Hémorragie cérébrale	(\bhip\b) ((hematom hemor)(?:.+)(\bft\b occip cerebral front parietal temporal cerebell fosse parench thalam cortical (intra.)?(ventric cranien)))) ((\bavc\b \bacv\b accident cerebro)(?:.+)\hemor)	431[.]
Hémorragie sous-arachnoidienne	(hematom hemor)(?:.+)\arachnoid \bhsa\b	430[.]90
Hépatite	hepatit	573[.]3 571[.]1
Hernie autre	hernie (inc omb)	55 1-3[.]
Hernie inguinale	hernie(s)? ing	550[.]
Hypertension	hypertensi \bhta\b	NA
Hypertension pulmonaire	((hypertension htn) pulm)\ bhtp\b	NA
ICT	(\bi(W)?c(W)?t\b) ((isch paresie)(?:.+)\transitoir) amauros(?:.+)\fuga	435[.]
Inféction urinaire	(\bpna\b) (\pyelonephrit) (p n a) (urosepsis) (\bcystite) (infec\w+ urinaire)	590[.]18 599[.]0 595[.]9 601
Infarctus de myocarde	(syn(?:.+)\bcoro) ((angor angine)(?:.+)\instable) (coronarien) (stemi) (s(W)?(t(W)?()?e(W)?()?m(W)?()?i) (\bsca\b) (\infarct(?:.+)\myocard) \infarctus (hemodynamie) (\bptca\b) coronaro	410[.]4 11[.]((9(^0))[(^9)])
Infection	(\b(sur)?infect)\ bfievre febrile hypertherm temperature	NA
Insuffisance artérielle	NA	443[.]9 440[.]9

Insuffisance cardiaque	(\bins\w* cardia)(\def\w+cardiaq)(\dysf\w+(?:+)(\bv\g\b ventricul diastol systol))(\bcmp\b)(\ic(?:+)(decompens diastol systol))\surchargel(oedem(?:+)(pulm poumon))(\boap\b)\anasarque	414[.][89][428[.][19][518[.][4812))][425[.][4]429[.][9]276[.][60]
Insuffisance hépatique	\bins(?:+)\hepatiq	573[.][81]
Insuffisance rénale	(\bi(\W)?r(\W)?a\b)(\bi(\W)?r(\W)?c\b)(\bins(?:*) renal)	58[1-3][.][584[.][90]586[.][9]
Insuffisance respiratoire	((\bins detresse difficulte)(?:+)\respirat)\bdesatur hypoxemi	786[.][01]786[.][0]b
Intoxication	(poly)?intox sevrage (dro[gq]ue medicam) toxicom abus drogue	977[.][291[.][30]345[.][.][292[.][0]
Ischémie artérielle	NA	459[.][932]
Ischémie intestinale	((enterite colite)(?:+)\ischémique)((ischémie angine)(?:+)(mesenter intestin))	557[.][09]
Lésion cérébrale	(masse lesion tumeur)(?:+)(cerebral frontal cerebell parietal occipital temporal)	NA
Lésion pancréatique	(masse tumeur lesion)(?:+)\pancrea(s tique)	NA
Lésion pulmonaire	(masse nodule lesion tumeur)(?:+)(poumon pulmon \b[si][gd]\b lobe (inf sup moy))\bbtt\b biopsie(?:+)\trans(?:+)\thorac	NA
Léucémie chronique	(synd\w+(?:+)(myelo lympho))(\waldenstrom)\bllc\b	20[45][.][1]
Leucémie aigue	leucemie\b ma\b lla\b	20[458][.][09]
Lithiase urinaire	calcul lithias colique neph corali form nephrostomie\b j\b	NA
Lymphome	lymphome\b lnh\b	202[.]
Mélanome	\bmelanom	172[.]
Maladie de Parkinson	parkinson	332[.][0]
Maladie diverticulaire	diverticul(it os ai)	562[.]
Maladie inflammatoire de l'intestin	(crohn)(colite ulc)	555[.][556[.][9]
Maladie vasculaire athérosclérotique (MVAS)	(ische(?:+)((\bgr)(membre)(pied))(mollet))(\claudic)((\bins(?:+))(\isch(?:+)))(\art)(\vascul)(\mi[dg]))(\amputat)\bfrein\b(\b(arterio angio)(\b graph plast))	443[.][9]440[.][9]459[.][932]785[.][4]
Meningite	meningit encephalit	32[0-4][.][.]
Metastase	metasta[st] carcinomatosel panchement(?:+)\neo\b bst(?:+)\b(4 iv)(.)?\b mesotheliom	19[789][.][.]
MPOC	(mpoc\b)(\malad\w+ pulm\w+ obstr\w+ chron\w+)(\m(\W)?(\p(\W)?(\o(\W)?(\c) emphysem	NA
Myélome multiple	(myelom\w+ multip)	203[.]
Névrалgie du trijumeau	neuralgie(?:+)\trijum	NA
Neutropénie fébrile	(neutro pancytopenie)(?:+)\febr fièvre(?:+)\neutro	288[.][02]
Ostéomyélite	osteomyelit(plaie pied)(?:+)\(infect diabet)	730[.][2]
Ostéoporose	osteoporo	733[.][01]
Péricardite	(myo peri)(?:*)cardite	423[.][91]
Pancréatite	pancreatite	577[.][0]
Perforation	perforation(?:+)\(intestin grele colon colique caec jeju duoden ilea)(\ulcere ulcus divertic chole appendic)\bperfore pneumoperit	NA
Perte d'autonomie	(pert\w+ ((d'autonomie)(\dautono\w+ autono\w+)))\deconditionnement epuisement	NA
Perte de poids ou fatigue	perte de poids(\bdeg\b)((\dim\w+)(\det\w+))(?:+)\(etat g)\faiblesse etat(?:+)\general\basthenie	783[.][2]780[.][7]
Pneumonie	(pneumonie)\empyeme	48[0123568][.][.]
Pneumonie d'aspiration	(aspiration)	507[.][0]
Pneumothorax	pneumothora\b bptx\b	512[.][.]
Polype intestinale	polyp(?:+)\(coliq\b caec sigmoid\b colo rect appendice\b grele\b intestin\b ileo)	211[.][3]
Procédure urologique	\brtuv\b \brtuv\b hbp \brtuv\b \bhp\b \baps\b prostate\b bhp\b hématurie	NA
Prolapsus	\bprolap(a)?sus (relachement relaxation)(?:+)\(pelvien)(\cysto recto colpo entero)(?:*)cele\b procidence\b colpo	618[.][788[.][3]
Prothèse articulaire	\bpth\b prot\w+ tot\w+	NA
Prothèse valvulaire	bentall tiron(?:+)\david(\brvm\b)(\brva\b)	NA
Réanimation	\bac\b arret cardi reanimation	427[.][5]798[.][1]
Régurgitation valvulaire	(\bins(?:+)\regurg(?:+)\(mitral aort)	424[.][01]1
Rétention urinaire	retention	788[.][2]
Sévère	\bsevere\b	NA
Saignement gastro-intestinale	\bhd[hb]\b rectorragie melena hémorr(?:+)\digest	578[.][569[.][31]
Sepsis	(sepsis)(\bseptiqu)	038[.][0-489][790[.][7]
Soins de fin de vie	(palliati) soins confort phase terminal	NA
Sténose carotidienne	st(?:+)\(carotid\b ci[gd]\b)\endarterect	433[.][12]

Sténose spinale	stenose(?:.+(foram spinal lombaire cervical cdtls d)) myelopathie laminect compression medul(l)?aire foraminect spondylolisthesi	NA
Sténose valvulaire	(\bstenos(?:.+(mitral aort bao\b))\b bsa\b	424[.]13 39 45[.]0
Syncope ou lipothymie	\bsyncop tension(?:.+)orthosta lipoth[yi]mi	780[.]2 458[.]0 458[.]9
Syndrome de Guillain-Barré	guil(?:.+)bar mill(?:.+)fis bcidp\b	357[.]0
Tachycardie	(\bt(\W)?s(\W)?v) (tachycardie) \bt(\W)?v\b arythmie(?:.+)ventric(\bt(\W)?a(\W)?p)	427[.]0-2
Tamponnade	tamponnade tamponade	NA
Trauma	trauma laceration plaie (luxation) (rupture)	95 89[.] 86 0-9[.]
Trauma cérébrale	(\bhsd\b) (hematom hemor)(?:.+(dural))\btcc\b (trauma contusion)(?:.+) (\bcranien cer ebra)	432[.]1 85 14[.]0
Tumeur	(tumeur) (nodule) (masse) (lesion) (adenome) meningiome adenopathie	NA
Tumeur benin	NA	2 12 [0-9][.]
TVP	emboli(?:.+)pulm \bt[vp]p\b t(h)?rombophleb thrombose veineuse phlebite embolie massive ep massive	415[.]1 451[.] 28
Varice oesophagienne	varice(?:.+)oeso hemorr(?:.+) (varic oeso)	456[.]1
Virus	\bsag\b (synd(?:.+) (gripp viral)) grippe grippal virus rhume influenz \bi(\W)?v(\W)?r(\W)?s\b	465[.]91 487[.]1 079[.]9

Annexe 2 Diagnostics antérieurs

Table de correspondance entre chaque variable booléenne (colonne « Variable ») dans l'ensemble « diagnostics antérieurs » incluse dans le modèle HEADS-UP et les codes CIM-10 dans les bases de données MED-ÉCHO et SIGDU. Les codes CIM-10 sont distingués selon leur provenance, soit de la codification du diagnostic représentant la raison d'admission suite à un séjour à l'urgence (colonne « Diagnostics à l'urgence »), de la codification du diagnostic principale sur la feuille sommaire (colonne « Diagnostics principales ») ou de la codification d'un diagnostic autre que principal sur la feuille sommaire (colonne « Diagnostics secondaires »). Fichier disponible sous forme numérique au format CSV sur demande (annexe_2_dx_antérieur.csv).

Variable	Diagnostics à l'urgence	Diagnostics secondaires	Diagnostics principales
a_fib	I48	I48	NA
acute_resp_failure	J960	J960	NA
alcohol	F10 E52 G621 I426 K292 Z721	F10 E52 G621 I426 K292 Z721	NA
anasarca	E877 R60	E877 R60	NA
anemia	NA	D5[1-3] D649 D638	NA
angina_recent	I20	NA	NA
anticoagulation	NA	Z921	NA
aortic_stenosis	NA	I350	NA
ascites	R18	R18	NA
asthma	J45	J45	NA
breast_cancer	NA	C50 Z853	NA
bronchiectasis	J47	J47	NA
cachexia	R64 R53	R64 R54	NA
cad	NA	I20 I25 Z955	NA
cancer_ed	C50 Z853 C767	NA	NA
chemo_cancer_1	NA	Z926	NA
chemo_cancer_2	Z511 D700	Z511	NA
chest_cancer_1	NA	Z85[12] C3[0-39]	NA
chest_cancer_2	C34	C3[4-8]	NA
chf	NA	I099 I11 I13 I255 I42[05679] I43 I50 P290	NA
chf_adm	I099 I11 I13 I255 I42[05679] I43 I50 P290	NA	I099 I11 I13 I255 I42[05679] I43 I50 P290
chronic_resp_failure	J96[19]	J96[19]	NA
cns_cancer	C71	C71	NA
cvd	G4[56] H340 I6[0-9]	G4[56] H340 I6[0-9]	NA
dementia	NA	F0[0-3] F051 G30 G31	NA
denutrition	NA	E4[36]	NA
depression	NA	F204 F31[345] F32 F33 F341 F412 F432	NA
dialysis	NA	Z992 Z49	NA
endo_1	NA	E1[0-4][09]	NA
endo_2	E1[0-4] G632 H360 N083	E1[0-4][1-8] G632 H360 N083	NA
endocrine_cancer	NA	C7[3-5]	NA
falls	R296 W06 W1[89] R268	R296 W06 W1[89]	NA
frailty	R64 R53 R296 W06 W1[89] R268	E4[36] R64 R54 R296 W06 W1[89]	NA
gi_cancer_1	NA	Z850 C1[789] C20	NA
gi_cancer_2	C18	C1[56] C2[4]	NA
gi_cancer_3	NA	C2[12356]	NA
gu_cancer_1	NA	C5[124] C560 C6[1467] Z85[45]	NA
gu_cancer_2	C57 C61	C5[3] C6[02358]	NA
gu_cancer_3	NA	C5[57] C56[1-9]	NA
heme_cancer_1	NA	C8[12358] Z85[67] C921 C9[16]	NA
heme_cancer_2	NA	C8[46] C9[04]	NA
heme_cancer_3	C8[1-9] C9[0-6]	C9[35] C92[04589]	NA
home_o2	NA	Z998	NA
ibd_crohn	K50	NA	K50
ild	NA	J84	NA
liver_1	NA	B18 K70[01239] K71[3457] K7[34] K759 K76[2-489] Z944 K86[01]	NA
liver_2	K7[1-6] Z944	K704 K711 K72 K758 K76[57]	NA
liver_rf	NA	I85[09] I864 I98[23] K766	NA
metastatic_solid_cancer	NA	C7[789] C80	NA
mi_past	NA	I21 I22 I252	NA

mi recent	I21 I22 I252	NA	NA
msk cancer	NA	C4[015-9]	NA
obesity hypoventilation	NA	E662	NA
obstetrics	O Z37	NA	O Z37
obstructive	J44[01] J43	J44 J43	NA
orl cancer	NA	C0[1-9] C1[0-4]	NA
palliative	NA	Z515	NA
paralysis	NA	G041 G114 G80[12] G8[12] G83[0-49]	NA
past pe	NA	NA	I26
pneumo adm	NA	NA	J4[0-467] J684 J70[13]
pneumonia adm	NA	NA	J1[2-8]
pseudomonas	NA	B965 J151	NA
psych	NA	F2[02-589] F302 F31[25]	NA
pulmonary hypertension	I27	I27	NA
pvd	I7[01] I73[189] I771 I79[02] K55[189] Z95[89]	I7[01] I73[189] I771 I79[02] K55[189] Z95[89]	NA
recent abdominal pain	R104	NA	NA
recent anemia	D649	NA	NA
recent back pain	M545	NA	NA
recent chest pain	R074	NA	NA
recent colitis	A099	NA	NA
recent complication	T8	NA	NA
recent gi bleed	K922	NA	NA
recent hip fracture	S72	NA	NA
recent interstitial nephritis	N10	NA	NA
recent intestinal occlusion	K56	NA	NA
recent pe	I26	NA	NA
recent perforation	K578	NA	NA
recent pneumonia	J1[2-8]	NA	NA
recent uti	N390	NA	NA
renal_1	N189	I12 I131 N03[2-7] N05[2-7] N18[1238] N19 Z940	NA
renal_2	NA	N18[045] N250	NA
skin cancer	C44	C4[34]	NA
transplant	NA	Z94[0-4]	NA
valve	NA	I0[5-8] I3[4-9] Z95[2-4]	NA

Annexe 3 Service d'admission

Les niveaux de la variable catégorique « Service d'admission » inclut dans les modèles HEADS-UP et mHOMR (colonne « Service d'admission ») avec les services normatifs correspondant dans la base de données MED-ÉCHO (colonne « Service(s) MED-ECHO correspondant(s) »). Fichier disponible sous forme numérique au format CSV sur demande (annexe_3_service.csv).

Service d'admission	Service(s) MED-ECHO correspondant(s)
Cardiologie	CARDIOLOGIE
Chirurgie cardiaque	CHIRURGIE CARDIAQUE, CHIRURGIE C.V.T
Chirurgie générale	CHIRURGIE GENERALE
Chirurgie maxillo-faciale	CHIR.BUCCALE/MAXILLO-FACIALE
Chirurgie orthopédique	ORTHOPEDIE
Chirurgie plastique	CHIRURGIE PLASTIQUE
Chirurgie thoracique	CHIR.THORACIQUE (POUMONS)
Chirurgie vasculaire	CHIRURGIE VASCULAIRE
Endocrinologie	ENDOCRINOLOGIE
Gastro-entérologie	GASTRO-ENTEROLOGIE
Gynécologie	GYNECOLOGIE
Hématologie / Oncologie	HEMATOLOGIE, ONCOLOGIE, RADIO-ONCOLOGIE
Médecine familiale	MEDECINE FAMILIALE
Médecine interne	MEDECINE INTERNE GERIATRIQUE, MEDECINE INTERNE, MICROBIOLOGIE MEDICALE, MEDECINE GENERALE GERIATRIQUE
Néphrologie	NEPHROLOGIE
Neurochirurgie	NEUROCHIRURGIE
Neurologie	NEUROLOGIE
Obstétrique	OBSTETRIQUE -- DELIVREES, OBSTETRIQUE -- NON-DELIVREES, OBSTETRIQUE -- AVORTEMENT, OBSTETRIQUE -- GYNECOLOGIE, OBSTETRIQUE -- POST-PARTUM
Ophthalmologie	OPHTALMOLOGIE
Oto-rhino-laryngologie	O.R.L.
Pneumologie	PNEUMOLOGIE
Rhumatologie	RHUMATOLOGIE
Soins intensifs	SOINS CRITIQUES, ANESTHESIE-REANIMATION
Soins palliatifs	SOINS PALLIATIFS
Trauma	TRAUMATOLOGIE
Urologie	UROLOGIE

Annexe 4 Regroupement personnalisé

Table de correspondance entre les variables booléennes dans notre regroupement des comorbidités (colonne « Variable ») et les codes CIM-10 rétrospectifs provenant des feuilles sommaires qui faisait partie de chaque groupe (colonne « Codes CIM-10 »). Le terme « personnalisé » réfère à la supposition que ce regroupement est plus approprié pour la prédiction de risque personnalisé que le regroupement du Charlson Comorbidity Index. Fichier disponible sous forme numérique au format CSV sur demande (annexe_4_dx_retro.csv).

Variable	Codes CIM-10
a fib	I48
acute respiratory failure	J960
anasarca	E877 R60
anemia	D5[1-3] D649 D638
anticoagulation	Z921
aortic stenosis	I350
ascites	R18
asthma	J45
breast cancer	C50 Z853
bronchiectasis	J47
bronchitis	J4[0-2]
cachexia	R64 R54
cad	I20 I25 Z955
chemo cancer 1	Z926
chemo cancer 2	Z511
chest cancer 1	Z85[12] C3[0-39]
chest cancer 2	C3[4-8]
chf	I099 I11 I13 I255 I42[05679] I43 I50 P290
chronic pulmonary disease	J4[0-467] J684 J70[13]
cns cancer	C71
copd	J44
cvd	G4[56] H340 I6[0-9]
dementia	F0[0-3] F051 G30 G31
denutrition	E4[36]
depression	F204 F31[345] F32 F33 F341 F412 F432
diabetes with chronic complication	E1[0-4][1-8] G632 H360 N083
diabetes without chronic complication	E1[0-4][09]
dialysis	Z992 Z49
emphysema	J43
endocrine cancer	C7[3-5]
falls	R296 W06 W1[89]
gi cancer 1	Z850 C1[789] C20
gi cancer 2	C1[56] C2[4]
gi cancer 3	C2[12356]
gi varices	I85[09] I864 I98[23]
gu cancer 1	C5[124] C560 C6[1467] Z85[45]
gu cancer 2	C5[3] C6[02358]
gu cancer 3	C5[57] C56[1-9]
heme cancer 1	C8[12358] Z85[67] C921 C9[16]
heme cancer 2	C8[46] C9[04]
heme cancer 3	C9[35] C92[04589]
hemiplegia or paraplegia	G041 G114 G80[12] G8[12] G83[0-49]
home oxygen	Z998
important liver disease	K704 K711 K72 K758 K76[57]
important renal disease	N18[045] N250
interstitial lung disease	J84
metastatic solid cancer	C7[789] C80
mild liver disease	B18 K70[01239] K71[3457] K7[34] K759 K76[2-489] Z944 K86[01]
mild renal disease	I12 I131 N03[2-7] N05[2-7] N18[1238] N19 Z940
msk cancer	C4[015-9]
myocardial infarction	I21 I22 I252
obesity hypoventilation	E662
orl cancer	C0[1-9] C1[0-4]
palliative	Z515
portal hypertension	K766

pseudomonas	B965 J151
psych	F2[02-589] F302 F31[25]
pulmonary hypertension	I27
pvd	I7[01] I73[189] I771 I79[02] K55[189] Z95[89]
respiratory failure	J96[19]
roh	F10 E52 G621 I426 K292 Z721
skin cancer	C4[34]
transplant	Z94[0-4]
valve	I0[5-8] I3[4-9] Z95[2-4]

Annexe 5 Coefficients de régression du modèle mHOMR

Coefficient de régression (colonne « Coefficient ») et la signification statistique (colonne « Valeur p ») de chaque variable (colonne « Variable ») dans le modèle mHOMR ajusté avec les données du CHUS. Fichier disponible sous forme numérique au format CSV sur demande (annexe_5_mhomr_coefficients.csv).

Variable	Coefficient	Valeur p
(Intercept)	-6.437659758	<0.001
Âge	0.791942407	<0.001
Sexe	0.272218409	<0.001
Service d'admission=Cardiologie	-1.643307396	<0.001
Service d'admission=Chirurgie cardiaque	-1.75750031	<0.001
Service d'admission=Chirurgie générale	-1.618468327	<0.001
Service d'admission=Chirurgie maxillo-faciale	-3.201127776	<0.001
Service d'admission=Chirurgie orthopédique	-2.548341251	<0.001
Service d'admission=Chirurgie plastique	-1.459421084	<0.001
Service d'admission=Chirurgie thoracique	-0.88299936	<0.001
Service d'admission=Chirurgie vasculaire	-1.565581336	<0.001
Service d'admission=Endocrinologie	-12.7689914	0.86
Service d'admission=Gastro-entérologie	-0.138760755	0.28
Service d'admission=Gynécologie	-1.999583996	<0.001
Service d'admission=Hématologie / Oncologie	0.640160379	<0.001
Service d'admission=Médecine familiale	0.037309059	0.82
Service d'admission=Néphrologie	-0.288449501	0.06
Service d'admission=Neurochirurgie	0.384071608	<0.001
Service d'admission=Neurologie	-0.442854962	0.00
Service d'admission=Obstétrique	-4.508793667	<0.001
Service d'admission=Ophtalmologie	-2.744833455	<0.001
Service d'admission=Oto-rhino-laryngologie	-1.463319396	<0.001
Service d'admission=Pneumologie	0.725234245	<0.001
Service d'admission=Rhumatologie	-0.904356475	0.01
Service d'admission=Soins intensifs	0.308650299	0.03
Service d'admission=Soins palliatifs	3.882033942	<0.001
Service d'admission=Trauma	-0.598185406	0.14
Service d'admission=Urologie	-2.186481337	<0.001
Urgence d'admission=Urgent sans ambulance	-0.951522993	<0.001
Urgence d'admission=Urgent avec ambulance	-0.600079734	0.00
Réadmission urgente=TRUE	0.546876854	<0.001
Visites à l'urgence	-0.75353174	<0.001
Admissions par ambulance	-0.489269701	0.00
Milieu de vie=Résidence	-0.53880358	0.08
Milieu de vie=Soins chroniques	0.636157724	0.08
Milieu de vie=Inconnu	-0.137397188	0.31
Admission aux soins intensifs=TRUE	0.690487225	<0.001
Service d'admission=Cardiologie:Urgence d'admission=Urgent sans ambulance	1.145953478	<0.001
Service d'admission=Chirurgie cardiaque:Urgence d'admission=Urgent sans ambulance	1.423000852	0.01
Service d'admission=Chirurgie générale:Urgence d'admission=Urgent sans ambulance	0.634771269	<0.001
Service d'admission=Chirurgie maxillo-faciale:Urgence d'admission=Urgent sans ambulance	2.366273242	0.06
Service d'admission=Chirurgie orthopédique:Urgence d'admission=Urgent sans ambulance	1.206692463	<0.001
Service d'admission=Chirurgie plastique:Urgence d'admission=Urgent sans ambulance	-0.48001455	0.24
Service d'admission=Chirurgie thoracique:Urgence d'admission=Urgent sans ambulance	0.992056836	0.01
Service d'admission=Chirurgie vasculaire:Urgence d'admission=Urgent sans ambulance	1.810130386	<0.001
Service d'admission=Endocrinologie:Urgence d'admission=Urgent sans ambulance	12.74835511	0.86
Service d'admission=Gastro-entérologie:Urgence d'admission=Urgent sans ambulance	0.712122908	<0.001
Service d'admission=Gynécologie:Urgence d'admission=Urgent sans ambulance	3.511758005	<0.001
Service d'admission=Hématologie / Oncologie:Urgence d'admission=Urgent sans ambulance	1.270294933	<0.001
Service d'admission=Médecine familiale:Urgence d'admission=Urgent sans ambulance	-0.159822598	0.39
Service d'admission=Néphrologie:Urgence d'admission=Urgent sans ambulance	0.774499633	0.00
Service d'admission=Neurochirurgie:Urgence d'admission=Urgent sans ambulance	1.058677744	<0.001
Service d'admission=Neurologie:Urgence d'admission=Urgent sans ambulance	0.006397587	0.98
Service d'admission=Obstétrique:Urgence d'admission=Urgent sans ambulance	2.93532587	0.01

Service d'admission=Ophtalmologie:Urgence d'admission=Urgent sans ambulance	1.534006402	0.01
Service d'admission=Oto-rhino-laryngologie:Urgence d'admission=Urgent sans ambulance	1.30075334	<0.001
Service d'admission=Pneumologie:Urgence d'admission=Urgent sans ambulance	0.908279128	<0.001
Service d'admission=Rhumatologie:Urgence d'admission=Urgent sans ambulance	-0.603882121	0.37
Service d'admission=Soins intensifs:Urgence d'admission=Urgent sans ambulance	0.179119144	0.61
Service d'admission=Soins palliatifs:Urgence d'admission=Urgent sans ambulance	12.7879862	0.93
Service d'admission=Trauma:Urgence d'admission=Urgent sans ambulance	-0.662670751	0.55
Service d'admission=Urologie:Urgence d'admission=Urgent sans ambulance	1.956762471	<0.001
Service d'admission=Cardiologie:Urgence d'admission=Urgent avec ambulance	1.122319831	<0.001
Service d'admission=Chirurgie cardiaque:Urgence d'admission=Urgent avec ambulance	2.743392032	<0.001
Service d'admission=Chirurgie générale:Urgence d'admission=Urgent avec ambulance	0.938901857	<0.001
Service d'admission=Chirurgie maxillo-faciale:Urgence d'admission=Urgent avec ambulance	2.010975662	0.12
Service d'admission=Chirurgie orthopédique:Urgence d'admission=Urgent avec ambulance	1.93259078	<0.001
Service d'admission=Chirurgie plastique:Urgence d'admission=Urgent avec ambulance	-0.1646291	0.73
Service d'admission=Chirurgie thoracique:Urgence d'admission=Urgent avec ambulance	0.190796485	0.54
Service d'admission=Chirurgie vasculaire:Urgence d'admission=Urgent avec ambulance	1.737633476	<0.001
Service d'admission=Endocrinologie:Urgence d'admission=Urgent avec ambulance	12.73659959	0.86
Service d'admission=Gastro-entérologie:Urgence d'admission=Urgent avec ambulance	0.127154073	0.41
Service d'admission=Gynécologie:Urgence d'admission=Urgent avec ambulance	2.677023323	<0.001
Service d'admission>Hématologie / Oncologie:Urgence d'admission=Urgent avec ambulance	1.236486668	<0.001
Service d'admission=Médecine familiale:Urgence d'admission=Urgent avec ambulance	-0.333456108	0.05
Service d'admission=Néphrologie:Urgence d'admission=Urgent avec ambulance	0.798326874	<0.001
Service d'admission=Neurochirurgie:Urgence d'admission=Urgent avec ambulance	0.456814932	<0.001
Service d'admission=Neurologie:Urgence d'admission=Urgent avec ambulance	0.325251501	0.05
Service d'admission=Obstétrique:Urgence d'admission=Urgent avec ambulance	-6.402597329	0.94
Service d'admission=Ophtalmologie:Urgence d'admission=Urgent avec ambulance	-9.87188944	0.96
Service d'admission=Oto-rhino-laryngologie:Urgence d'admission=Urgent avec ambulance	1.473012174	<0.001
Service d'admission=Pneumologie:Urgence d'admission=Urgent avec ambulance	0.320119042	0.01
Service d'admission=Rhumatologie:Urgence d'admission=Urgent avec ambulance	-0.142255346	0.75
Service d'admission=Soins intensifs:Urgence d'admission=Urgent avec ambulance	0.539357193	0.00
Service d'admission=Soins palliatifs:Urgence d'admission=Urgent avec ambulance	0.854211265	0.23
Service d'admission=Trauma:Urgence d'admission=Urgent avec ambulance	0.114708292	0.79
Service d'admission=Urologie:Urgence d'admission=Urgent avec ambulance	2.149957943	<0.001
Admissions par ambulance:Milieu de vie=Résidence	0.319020498	0.46
Admissions par ambulance:Milieu de vie=Soins chroniques	-0.381063877	0.40
Admissions par ambulance:Milieu de vie=Inconnu	-0.06168015	0.66
Urgence d'admission=Urgent sans ambulance:Admissions par ambulance	-0.030517747	0.89
Urgence d'admission=Urgent avec ambulance:Admissions par ambulance	0.251289645	0.14

Annexe 6 Importance des variables

L'importance des variables dans le modèle HEADS-UP. Modèle ajusté avec l'échantillon d'entraînement et l'importance calculée avec les cas « out-of-bag ». Seul le quart des variables les plus importantes sont affichés. La spécification des variables diagnostiques est dans l'Annexe 1 et 2.



Références

- [1] T. Sinuff *et al.*, “Improving End-of-Life Communication and Decision Making: The Development of a Conceptual Framework and Quality Indicators,” *J. Pain Symptom Manage.*, vol. 49, no. 6, pp. 1070–1080, Jun. 2015, doi: 10.1016/j.jpainsymman.2014.12.007.
- [2] D. K. Heyland *et al.*, “Failure to engage hospitalized elderly patients and their families in advance care planning,” *JAMA Intern. Med.*, vol. 173, no. 9, pp. 778–787, May 2013, doi: 10.1001/jamainternmed.2013.180.
- [3] A. Brinkman-Stoppelenburg, J. A. C. Rietjens, and A. van der Heide, “The effects of advance care planning on end-of-life care: A systematic review,” *Palliat. Med.*, vol. 28, no. 8, pp. 1000–1025, 2014, doi: 10.1177/0269216314526272.
- [4] J. J. You *et al.*, “Barriers to goals of care discussions with seriously ill hospitalized patients and their families: a multicenter survey of clinicians,” *JAMA Intern. Med.*, vol. 175, no. 4, pp. 549–556, Apr. 2015, doi: 10.1001/jamainternmed.2014.7732.
- [5] S. Lund, A. Richardson, and C. May, “Barriers to advance care planning at the end of life: an explanatory systematic review of implementation studies,” *PloS One*, vol. 10, no. 2, p. e0116629, Jan. 2015, doi: 10.1371/journal.pone.0116629.
- [6] H. J. Warraich, L. A. Allen, K. J. Mukamal, A. Ship, and R. D. Kociol, “Accuracy of physician prognosis in heart failure and lung cancer: Comparison between physician estimates and model predicted survival,” *Palliat. Med.*, vol. 30, no. 7, pp. 684–689, 2016, doi: 10.1177/0269216315626048.
- [7] A. De Vleminck *et al.*, “Barriers to Advance Care Planning in Cancer, Heart Failure and Dementia Patients: A Focus Group Study on General Practitioners’ Views and Experiences,” *PLoS ONE*, vol. 9, no. 1, Jan. 2014, doi: 10.1371/journal.pone.0084905.
- [8] N. A. Christakis and E. B. Lamont, “Extent and determinants of error in doctors’ prognoses in terminally ill patients: prospective cohort study,” *BMJ*, vol. 320, no. 7233, pp. 469–472, Feb. 2000, doi: 10.1136/bmj.320.7233.469.
- [9] K. G. M. Moons, P. Royston, Y. Vergouwe, D. E. Grobbee, and D. G. Altman, “Prognosis and prognostic research: what, why, and how?,” *BMJ*, vol. 338, p. b375, Feb. 2009, doi: 10.1136/bmj.b375.
- [10] K. G. M. Moons *et al.*, “Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD): Explanation and Elaboration,” *Ann. Intern. Med.*, vol. 162, no. 1, p. W1, Jan. 2015, doi: 10.7326/M14-0698.
- [11] E. W. Steyerberg, *Clinical Prediction Models: A Practical Approach to Development, Validation, and Updating*, 2nd ed. 2019 edition. S.l.: Springer, 2020.
- [12] L. C. Yourman, S. J. Lee, M. A. Schonberg, E. W. Widera, and A. K. Smith, “Prognostic Indices for Older Adults: A Systematic Review,” *JAMA*, vol. 307, no. 2, pp. 182–192, Jan. 2012, doi: 10.1001/jama.2011.1966.
- [13] E. W. Steyerberg *et al.*, “Prognosis Research Strategy (PROGRESS) 3: Prognostic Model Research,” *PLoS Med.*, vol. 10, no. 2, Feb. 2013, doi: 10.1371/journal.pmed.1001381.
- [14] H. Hemingway *et al.*, “Prognosis research strategy (PROGRESS) 1: A framework for researching clinical outcomes,” *BMJ*, vol. 346, p. e5595, Feb. 2013, doi: 10.1136/bmj.e5595.
- [15] A. Rajkomar, J. Dean, and I. Kohane, “Machine Learning in Medicine,” *N. Engl. J. Med.*, vol. 380, no. 14, pp. 1347–1358, Apr. 2019, doi: 10.1056/NEJMr1814259.
- [16] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd edition. Springer, 2009.

- [17] L. Breiman, “Statistical Modeling: The Two Cultures (with comments and a rejoinder by the author),” *Stat. Sci.*, vol. 16, no. 3, pp. 199–231, Aug. 2001, doi: 10.1214/ss/1009213726.
- [18] E. Steyerberg, *Clinical Prediction Models: A Practical Approach to Development, Validation, and Updating*. Springer Science & Business Media, 2008.
- [19] K. G. M. Moons, D. G. Altman, Y. Vergouwe, and P. Royston, “Prognosis and prognostic research: application and impact of prognostic models in clinical practice,” *BMJ*, vol. 338, Jun. 2009, doi: c.
- [20] B. Van Calster, D. Nieboer, Y. Vergouwe, B. De Cock, M. J. Pencina, and E. W. Steyerberg, “A calibration hierarchy for risk models was defined: from utopia to empirical data,” *J. Clin. Epidemiol.*, vol. 74, pp. 167–176, Jun. 2016, doi: 10.1016/j.jclinepi.2015.12.005.
- [21] M. A. Ahmad, C. Eckert, A. Teredesai, and G. McKelvey, “Interpretable Machine Learning in Healthcare,” *IEEE Intell. Inform. Bull.*, vol. 19, no. 1, p. 7, Aug. 2018.
- [22] Z. C. Lipton, “The Mythos of Model Interpretability,” *ArXiv160603490 Cs Stat*, Jun. 2016, Accessed: Aug. 26, 2019. [Online]. Available: <http://arxiv.org/abs/1606.03490>.
- [23] F. Doshi-Velez and B. Kim, “Towards A Rigorous Science of Interpretable Machine Learning,” *ArXiv170208608 Cs Stat*, Feb. 2017, Accessed: Aug. 27, 2019. [Online]. Available: <http://arxiv.org/abs/1702.08608>.
- [24] E. W. Steyerberg, “Evaluation of Clinical Usefulness,” *Clin. Predict. Models*, pp. 309–328, 2019, doi: 10.1007/978-3-030-16399-0_16.
- [25] H. F. Ryder, C. McDonough, A. N. A. Tosteson, and J. D. Lurie, “Decision Analysis and Cost-effectiveness Analysis,” *Semin. Spine Surg.*, vol. 21, no. 4, pp. 216–222, Dec. 2009, doi: 10.1053/j.semss.2009.08.003.
- [26] S. G. Baker, N. R. Cook, A. Vickers, and B. S. Kramer, “Using relative utility curves to evaluate risk prediction,” *J. R. Stat. Soc. Ser. A Stat. Soc.*, vol. 172, no. 4, pp. 729–748, Oct. 2009, doi: 10.1111/j.1467-985X.2009.00592.x.
- [27] A. J. Vickers and E. B. Elkin, “Decision curve analysis: a novel method for evaluating prediction models,” *Med. Decis. Mak. Int. J. Soc. Med. Decis. Mak.*, vol. 26, no. 6, pp. 565–574, Dec. 2006, doi: 10.1177/0272989X06295361.
- [28] B. Van Calster *et al.*, “Reporting and Interpreting Decision Curve Analysis: A Guide for Investigators,” *Eur. Urol.*, vol. 74, no. 6, pp. 796–804, 2018, doi: 10.1016/j.eururo.2018.08.038.
- [29] X. Liu *et al.*, “A comparison of deep learning performance against health-care professionals in detecting diseases from medical imaging: a systematic review and meta-analysis,” *Lancet Digit. Health*, vol. 1, no. 6, pp. e271–e297, Oct. 2019, doi: 10.1016/S2589-7500(19)30123-2.
- [30] A. Avati, K. Jung, S. Harman, L. Downing, A. Ng, and N. H. Shah, “Improving palliative care with deep learning,” *BMC Med. Inform. Decis. Mak.*, vol. 18, no. Suppl 4, p. 122, 12 2018, doi: 10.1186/s12911-018-0677-8.
- [31] C. van Walraven, F. A. McAlister, J. A. Bakal, S. Hawken, and J. Donzé, “External validation of the Hospital-patient One-year Mortality Risk (HOMR) model for predicting death within 1 year after hospital admission,” *CMAJ*, vol. 187, no. 10, pp. 725–733, Jul. 2015, doi: 10.1503/cmaj.150209.
- [32] C. van Walraven, “The Hospital-patient One-year Mortality Risk score accurately predicted long-term death risk in hospitalized patients,” *J. Clin. Epidemiol.*, vol. 67, no. 9, pp. 1025–1034, Sep. 2014, doi: 10.1016/j.jclinepi.2014.05.003.

- [33] C. van Walraven and A. J. Forster, “The HOMR-Now! Model Accurately Predicts 1-Year Death Risk for Hospitalized Patients on Admission,” *Am. J. Med.*, Mar. 2017, doi: 10.1016/j.amjmed.2017.03.008.
- [34] P. Wegier *et al.*, “mHOMR: a feasibility study of an automated system for identifying inpatients having an elevated risk of 1-year mortality,” *BMJ Qual. Saf.*, p. bmjqs-2018-009285, Jun. 2019, doi: 10.1136/bmjqs-2018-009285.
- [35] E. W. Steyerberg *et al.*, “Assessing the performance of prediction models: a framework for some traditional and novel measures,” *Epidemiol. Camb. Mass*, vol. 21, no. 1, pp. 128–138, Jan. 2010, doi: 10.1097/EDE.0b013e3181c30fb2.
- [36] G. James, D. Witten, T. Hastie, and R. Tibshirani, “Linear Model Selection and Regularization,” in *An Introduction to Statistical Learning: with Applications in R*, G. James, D. Witten, T. Hastie, and R. Tibshirani, Eds. New York, NY: Springer, 2013, pp. 203–264.
- [37] F. E. Harrell, “Multivariable Modeling Strategies,” in *Regression Modeling Strategies: With Applications to Linear Models, Logistic and Ordinal Regression, and Survival Analysis*, Jr. Harrell Frank E., Ed. Cham: Springer International Publishing, 2015, pp. 63–102.
- [38] P. Powers, “Québec’s e-Health Journey – Assessing the Path Forward,” *ElectronicHealthcare*, Jan. 02, 2012. <https://www.longwoods.com/content/22745> (accessed Aug. 28, 2019).
- [39] N. Sahni, G. Simon, and R. Arora, “Development and Validation of Machine Learning Models for Prediction of 1-Year Mortality Utilizing Electronic Medical Record Data Available at the End of Hospitalization in Multiccondition Patients: a Proof-of-Concept Study,” *J. Gen. Intern. Med.*, vol. 33, no. 6, pp. 921–928, Jun. 2018, doi: 10.1007/s11606-018-4316-y.
- [40] L. Wang *et al.*, “Development and Validation of a Deep Learning Algorithm for Mortality Prediction in Selecting Patients With Dementia for Earlier Palliative Care Interventions,” *JAMA Netw. Open*, vol. 2, no. 7, pp. e196972–e196972, Jul. 2019, doi: 10.1001/jamanetworkopen.2019.6972.
- [41] A. J. Vickers, B. V. Calster, and E. W. Steyerberg, “Net benefit approaches to the evaluation of prediction models, molecular markers, and diagnostic tests,” *BMJ*, vol. 352, p. i6, Jan. 2016, doi: 10.1136/bmj.i6.
- [42] T. H. Kappen, W. A. van Klei, L. van Wolfswinkel, C. J. Kalkman, Y. Vergouwe, and K. G. M. Moons, “Evaluating the impact of prediction models: lessons learned, challenges, and recommendations,” *Diagn. Progn. Res.*, vol. 2, no. 1, p. 11, Jun. 2018, doi: 10.1186/s41512-018-0033-6.
- [43] “InfoCentre.” <https://www.crchus.ca/services-outils/autres-services-et-outils/infocentre/> (accessed Jul. 12, 2020).
- [44] A. Grant *et al.*, “Integrating feedback from a clinical data warehouse into practice organisation,” *Int. J. Med. Inf.*, vol. 75, no. 3–4, pp. 232–239, Apr. 2006, doi: 10.1016/j.ijmedinf.2005.07.037.
- [45] “MED-ECHO - Sources de données et métadonnées - Professionnels de la santé - MSSS.” <https://www.msss.gouv.qc.ca/professionnels/documentation-sources-de-donnees-et-indicateurs/sources-de-donnees-et-metadonnees/med-echo/> (accessed Jul. 12, 2020).
- [46] “SIGDU - Système d’information de gestion des urgences - Actifs informationnels - Professionnels de la santé - MSSS.” <https://www.msss.gouv.qc.ca/professionnels/technologies-de-l-information/actifs-informationnels/sigdu/> (accessed Jul. 12, 2020).

- [47] L. Breiman, “Random Forests,” *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, Oct. 2001, doi: 10.1023/A:1010933404324.
- [48] G. James, D. Witten, T. Hastie, and R. Tibshirani, *An Introduction to Statistical Learning: with Applications in R*, 1st ed. 2013, Corr. 7th printing 2017 edition. New York: Springer/Sci-Tech/Trade, 2013.
- [49] R. B. Parikh *et al.*, “Machine Learning Approaches to Predict 6-Month Mortality Among Patients With Cancer,” *JAMA Netw. Open*, vol. 2, no. 10, Oct. 2019, doi: 10.1001/jamanetworkopen.2019.15997.
- [50] C. Strobl, A.-L. Boulesteix, A. Zeileis, and T. Hothorn, “Bias in random forest variable importance measures: Illustrations, sources and a solution,” *BMC Bioinformatics*, vol. 8, no. 1, p. 25, Jan. 2007, doi: 10.1186/1471-2105-8-25.
- [51] M. N. Wright and A. Ziegler, “ranger: A Fast Implementation of Random Forests for High Dimensional Data in C++ and R,” *J. Stat. Softw.*, vol. 77, no. 1, 2017, doi: 10.18637/jss.v077.i01.
- [52] T. P. A. Debray, Y. Vergouwe, H. Koffijberg, D. Nieboer, E. W. Steyerberg, and K. G. M. Moons, “A new framework to enhance the interpretation of external validation studies of clinical prediction models,” *J. Clin. Epidemiol.*, vol. 68, no. 3, pp. 279–289, Mar. 2015, doi: 10.1016/j.jclinepi.2014.06.018.
- [53] J. D. Malley, J. Kruppa, A. Dasgupta, K. G. Malley, and A. Ziegler, “Probability Machines: Consistent Probability Estimation Using Nonparametric Learning Machines,” *Methods Inf. Med.*, vol. 51, no. 1, pp. 74–81, Jan. 2012, doi: 10.3414/ME00-01-0052.
- [54] C. A. Field and A. H. Welsh, “Bootstrapping clustered data,” *J. R. Stat. Soc. Ser. B Stat. Methodol.*, vol. 69, no. 3, pp. 369–390, 2007, doi: 10.1111/j.1467-9868.2007.00593.x.
- [55] Allen Larry A. *et al.*, “Decision Making in Advanced Heart Failure,” *Circulation*, vol. 125, no. 15, pp. 1928–1952, Apr. 2012, doi: 10.1161/CIR.0b013e31824f2173.
- [56] L. M. Chihara and T. C. Hesterberg, *Mathematical Statistics with Resampling and R*. John Wiley & Sons, 2018.
- [57] “PARS3 • Solutions • GRIIS,” *GRIIS*. <https://griis.ca/en/solutions/pars3/> (accessed Jul. 12, 2020).
- [58] D. Curtin *et al.*, “Predicting 1-Year Mortality in Older Hospitalized Patients: External Validation of the HOMR Model,” *J. Am. Geriatr. Soc.*, vol. 67, no. 7, pp. 1478–1483, Jul. 2019, doi: 10.1111/jgs.15958.
- [59] N. Rieke *et al.*, “The future of digital health with federated learning,” *NPJ Digit. Med.*, vol. 3, Sep. 2020, doi: 10.1038/s41746-020-00323-1.
- [60] J. Futoma, M. Simons, T. Panch, F. Doshi-Velez, and L. A. Celi, “The myth of generalisability in clinical research and machine learning in health care,” *Lancet Digit. Health*, vol. 2, no. 9, pp. e489–e492, Sep. 2020, doi: 10.1016/S2589-7500(20)30186-2.
- [61] R. E. Bernacki, S. D. Block, and American College of Physicians High Value Care Task Force, “Communication about serious illness care goals: a review and synthesis of best practices,” *JAMA Intern. Med.*, vol. 174, no. 12, pp. 1994–2003, Dec. 2014, doi: 10.1001/jamainternmed.2014.5271.
- [62] S. Gupta *et al.*, “Position statement from the Canadian Thoracic Society (CTS) on clinical triage thresholds in respiratory disease patients in the event of a major surge during the COVID-19 pandemic,” *Can. J. Respir. Crit. Care Sleep Med.*, vol. 4, no. 4, pp. 214–225, Oct. 2020, doi: 10.1080/24745332.2020.1769436.

- [63] T. Baker and M. Gerdin, "The clinical usefulness of prognostic prediction models in critical illness," *Eur. J. Intern. Med.*, vol. 45, pp. 37–40, Nov. 2017, doi: 10.1016/j.ejim.2017.09.012.
- [64] H. Singh, C. Spitzmueller, N. J. Petersen, M. K. Sawhney, and D. F. Sittig, "Information Overload and Missed Test Results in Electronic Health Record–Based Settings," *JAMA Intern. Med.*, vol. 173, no. 8, p. 702, Apr. 2013, doi: 10.1001/2013.jamainternmed.61.
- [65] L. F. Laker, C. M. Froehle, J. B. Windeler, and C. J. Lindsell, "Quality and Efficiency of the Clinical Decision-Making Process: Information Overload and Emphasis Framing," *Prod. Oper. Manag.*, vol. 27, no. 12, pp. 2213–2225, 2018, doi: <https://doi.org/10.1111/poms.12777>.
- [66] T. Shah, S. Patel-Teague, L. Kroupa, A. N. D. Meyer, and H. Singh, "Impact of a national QI programme on reducing electronic health record notifications to clinicians," *BMJ Qual. Saf.*, vol. 28, no. 1, pp. 10–14, Jan. 2019, doi: 10.1136/bmjqs-2017-007447.
- [67] M. J. Kissler, K. Kissler, and M. Burden, "Toward a Medical 'Ecology of Attention,'" *N. Engl. J. Med.*, Jan. 2021, doi: 10.1056/NEJMp2027190.
- [68] S. Bacchi, L. Oakden-Rayner, T. Zerner, T. Kleinig, S. Patel, and J. Jannes, "Deep Learning Natural Language Processing Successfully Predicts the Cerebrovascular Cause of Transient Ischemic Attack-Like Presentations," *Stroke*, vol. 50, no. 3, pp. 758–760, Mar. 2019, doi: 10.1161/STROKEAHA.118.024124.
- [69] A. Rajkomar *et al.*, "Scalable and accurate deep learning with electronic health records," *Npj Digit. Med.*, vol. 1, no. 1, pp. 1–10, May 2018, doi: 10.1038/s41746-018-0029-1.
- [70] B. J. Marafino *et al.*, "Validation of Prediction Models for Critical Care Outcomes Using Natural Language Processing of Electronic Health Record Data," *JAMA Netw. Open*, vol. 1, no. 8, p. e185097, Dec. 2018, doi: 10.1001/jamanetworkopen.2018.5097.
- [71] A. Névél, H. Dalianis, S. Velupillai, G. Savova, and P. Zweigenbaum, "Clinical Natural Language Processing in languages other than English: opportunities and challenges," *J. Biomed. Semant.*, vol. 9, no. 1, p. 12, Mar. 2018, doi: 10.1186/s13326-018-0179-8.
- [72] H. Quan *et al.*, "Coding Algorithms for Defining Comorbidities in ICD-9-CM and ICD-10 Administrative Data," *Med. Care*, vol. 43, no. 11, pp. 1130–1139, 2005.
- [73] S. Varma and R. Simon, "Bias in error estimation when using cross-validation for model selection," *BMC Bioinformatics*, vol. 7, p. 91, Feb. 2006, doi: 10.1186/1471-2105-7-91.
- [74] A. M. Molinaro, R. Simon, and R. M. Pfeiffer, "Prediction error estimation: a comparison of resampling methods," *Bioinforma. Oxf. Engl.*, vol. 21, no. 15, pp. 3301–3307, Aug. 2005, doi: 10.1093/bioinformatics/bti499.